Genetic variation in New Zealand abalone, *Haliotis iris*

A thesis submitted in partial fulfillment

of the of the requirements for the Degree of

Doctor of Philosophy in Biological Sciences

at the University of Canterbury

by Margaret Will

University of Canterbury
Christchurch, New Zealand

2009

Table of Contents

ABSTRACT 1

1. INTRODUCTION 3

GENE FLOW 3
ABALONE 7
Systematics 9
Assessing genetic structure 11
Genetic structure of abalone 12
NEW ZEALAND ABALONE 20
Aims 22

2. GENETIC STRUCTURE ACROSS COOK STRAIT 25

ABSTRACT 25
INTRODUCTION 25
Genetic splits around Cook Strait 26
Cook Strait as a barrier 28
H. iris genetic structure 31
MATERIALS AND METHODS 32
Marker choice 32
Samples 33
DNA extraction, PCR amplification, and sequencing 33
Population genetic analyses 34
RESULTS 40
Haplotypes: Relationships and distributions 41
Cluster analyses 41
A priori hypothesis testing 42
Nested clade phylogeographic analysis 43
Mismatch distributions and neutrality tests 44
DISCUSSION 44
Chatham Islands vs. North and South Island 45
North Island vs. South Island 47
Sample sizes 50
Future work 51

3. ISOLATION AND CHARACTERIZATION OF *HALIOTIS IRIS* MICROSATELLITES 53

ABSTRACT 53
INTRODUCTION 53
Haliotis and conserved microsatellites 54
H. iris and microsatellites 55
MATERIALS AND METHODS 56
Cross-species amplification 56
Microsatellite development and screening 57
Allele scoring 59
Genotyping error rate 59
Locus characterizations 60
RESULTS 60
Cross-species amplification 60
Screening *H. iris* loci 60

Genotyping error rate 61
Characterization of AB14, AB21, and AB31 61
DISCUSSION 62
Cross-species amplification 62
H. iris microsatellites 63
Ongoing projects 67
4. POPULATION GENETIC STRUCTURE BASED ON THREE MICROSATELLITE LOCI 69
ABSTRACT 69
INTRODUCTION 69
Molecular markers and the Cook Strait 70
MATERIALS AND METHODS 71
Population genetic analyses 71
RESULTS 73
Genetic structure 74
DISCUSSION 74
Inbreeding and abalone 77
5. VARIATION IN GAMETE RECOGNITION PROTEINS 79
ABSTRACT 79
INTRODUCTION 79
Abalone lysin 81
MATERIALS AND METHODS 83
Haplotype inference 84
Population genetic analyses 84
RESULTS 85
DISCUSSION 86
Lysin exon variation 87
Lysin intron variation 90
Population genetic structure 90
6. VARIATION IN AN INTRON OF THE Gα1 PROTEIN AND ITS UTILITY FOR INFERRING EVOLUTIONARY PROCESSES IN LYSIN 93
ABSTRACT 93
INTRODUCTION 93
Testing selection 93
MATERIALS AND METHODS 96
Sequence verification 97
Haplotype inference 97
Population genetic analyses 98
RESULTS 99
DISCUSSION 101
Comparison with lysin 101
Genetic structure 102
7. SUMMARY AND CONCLUSIONS 105
Cook Strait 105
Lysin 107
Genetic variation in *H. iris* 108
Potential problems and future investigations 109

TABLES AND FIGURES 112
ACKNOWLEDGEMENTS 178
LITERATURE CITED 179
APPENDICES 205
APPENDIX 1: SAMPLES 206
APPENDIX 2: MITOCHONDRIAL HAPLOTYPES 207
APPENDIX 3: MITOCHONDRIAL HAPLOTYPE FREQUENCIES 241
APPENDIX 4: MISMATCH DISTRIBUTIONS 245
APPENDIX 6: *HALIOTIS IRIS* MICROSATELLITE LOCI 293
APPENDIX 7: *HALIOTIS IRIS* MICROSATELLITE ALLELE FREQUENCIES 297
APPENDIX 8: LYSIN HAPLOTYPES 302
APPENDIX 9: LYSIN GENOTYPE FREQUENCIES 308
APPENDIX 10: Gα1 INTRON HAPLOTYPES 310
APPENDIX 11: Gα1 INTRON GENOTYPE FREQUENCIES 336

Abstract

Abalone (*Haliotis* spp.) are marine broadcast spawners that inhabit temperate and tropical waters across the globe. Their importance as a fisheries resource has resulted in considerable research into key aspects of their biology, particularly around growth and reproduction. In addition, there has been ongoing interest regarding the genetic variation in both wild and hatchery populations. The majority of abalone dispersal probably occurs during a pelagic lecithotrophic larval stage. In general, oceanographic features, life history characteristics, and larval dispersal ability can manipulate dispersal and gene flow patterns of marine fauna. In the case of abalone, considerable research has examined the population genetic structure of a variety of species, and several papers implicate ocean currents and life history characteristics as important factors that define population genetic structure. In comparison to other abalone species, little information regarding the genetic structure of New Zealand's endemic *H. iris* exists. The goal of this thesis was to elucidate the genetic structure of *H. iris* using mitochondrial and nuclear markers in regards to two potential barriers to gene flow, the Cook Strait region and the gamete recognition protein, lysin.

The genetic structure of *H. iris* was first examined in regards to a consistent pattern of genetic structure emerging in recent literature of coastal marine invertebrates around New Zealand: specifically, a north-south genetic split that occurs in the Cook Strait region (Chapter 2). Two regions of the mitochondria (totaling 1055 bp) were amplified across 477 individuals from 25 locations around New Zealand. A north-south split around the Cook Strait region was evident among *H. iris* samples. Unlike the other studies of New Zealand coastal marine invertebrates, the north-south split for *H. iris* was not located across regions of reported upwelling; instead the split was located across Cook Strait narrows. The north-south split was reflected in increased haplotype diversity for the northern samples.

Genetic structure was also examined using microsatellite loci. After unsuccessful attempts at cross-species amplification using 8 loci developed for *H. rubra* and 11 loci developed for *H. midae*, 13 polymorphic loci were isolated de novo for *H. iris* (Chapter 3). Of these, three very polymorphic loci were optimized for population genetic analyses. These three loci were used to genotype 447–459 individuals from the same 25 locations examined with mitochondrial DNA (Chapter 4). Like the mitochondrial DNA, the microsattelites indicated population genetic structure around the Cook Strait region; however the split identified with microsatellites occurred in the greater Cook Strait region with two sample sites from the north of the South Island grouping with the North Island.

Extrinsic barriers, like the Cook Strait region, are the primary focus of studies examining differentiation in marine invertebrate fauna. However, dispersal of an individual to a new population does not necessarily mean it can successfully reproduce with individuals of the new population. Potentially, populations may be diverging at genes essential for reproduction, i.e. gamete recognition proteins. The abalone egg recognition protein, lysin, is one of the best characterized gamete recognition proteins in marine broadcast spawners. Despite its well-understood function and structure, studies examining variation in lysin have been limited to small sample sizes ($N \leq 11$) and have found very little variation. Here, lysin was screened across 287 individuals from 17 sampling sites around New Zealand to assess intraspecific variation and genetic structure across the Cook Strait region (Chapter 5). The majority of the variation in a 783 bp fragment spanning from exon 4 to 5, was located in the intron. The variability in this fragment detected no genetic structure among samples or across the Cook Strait region.

The variation in lysin was considerably lower than the variation in either the mitochondrial DNA or the microsatellite loci. To determine whether this was an artifact of being a nuclear sequence, which, in general, have a lower mutation rates than microsatellite markers and mitochondrial DNA and a larger effective population size then mitochondrial DNA, or was a signature of a recent selective sweep, 857 bp of the Gα1 intron was assessed for genetic variation in 227 *H. iris* individuals from 14 sampling locations (Chapter 6). The Gα1 intron was considerably more diverse than the lysin fragment examined, suggesting that the relative lack of variation in the lysin fragment has resulted from a recent selective sweep. Additionally, the Gα1 intron was used to examine population genetic structure across the Cook Strait region and detected a weak but significant pattern of structure consistent with that detected using the microsatellite loci.

Overall, the a priori tests of genetic structure based on mitochondrial DNA, microsatellite markers, and the across Gα1 intron all identified a north-south genetic split around the Cook Strait region; however, the patterns of this split was slightly inconsistent among molecular markers. When cluster analyses were applied the patterns of genetic structure became more similar: for the mitochondrial, microsatellite, and Gα1 intron data, cluster analyses indicate that only one sample from the north of the South Island groups with the North Island, while a few discrepancies existed in regards the grouping of samples from the east coast of the North Island.

1. Introduction

The phylum Mollusca is amazingly diverse with species estimates ranging from around 50,000 to 128,000 (Powell 1979). Given the marine origin and vast distribution of many molluscs, mollusks highlight a central conundrum for marine, population, and evolutionary biologists: How does differentiation occur in the world's oceans? For many marine taxa with bipartite life histories this question is particularly taxing. Superficially, the ocean is a massive continuous medium, and simply the presence of pelagic larvae suggests high levels of gene flow: fertilized eggs and subsequent developmental stages are released and seemingly dispersed among populations throughout the ocean. As a result, dispersed individuals would maintain genetic continuity between populations. Ultimately this scenario predicts that genetic divergence between populations would be minor and high levels of diversity would be the ancient products of slow divergence. On the contrary due to both extrinsic and intrinsic factors, many marine populations are divergent, and many evolutionary radiations are fast (Palumbi 1992).

Abalone (*Haliotis* spp.) are marine broadcast spawners with bipartite life histories. Their commercial importance has led to numerous intraspecifc studies in different species. In general, these species are characterized by little genetic structure and high levels of intraspecific variation. However, a few examples of marked genetic divergence occur. Previous studies of New Zealand's *H. iris* suggest panmixia; however, these studies were limited in the sampling locations used and the molecular markers employed. In contrast, studies of other New Zealand coastal marine invertebrates indicate genetic structure with populations divided around the Cook Strait region. This thesis seeks to answer:

1. Does genetic structure exist in *H. iris*?
2. Is the pattern of structure in *H. iris* consistent with other New Zealand marine invertebrates?
3. Does the pattern correspond with previously established barriers to gene flow?

Is the pattern of genetic structure concordant across molecular markers?

GENE FLOW

Evolutionary theory is founded on changes that occur at the level of populations. If population differentiation is a necessary precursor to speciation, then through understanding barriers to gene flow at the population level, we can deliberate on how population

differentiation and, potentially, speciation occurs. We can extrapolate present day mechanisms to predict the consequences of future environmental change (Underwood and Fairweather 1989) and to interpret the distribution, composition, and radiations of species documented in the fossil record (Jablonski 1986). If differentiation results from a lack of gene flow between populations, then identifying patterns of differentiation and putative barriers to gene flow are essential for understanding evolution.

Gene flow is the transfer of genes (or the migration of individuals) between populations. Molecular techniques are frequently employed to determine population genetic structure, and from this structure, barriers to gene flow are proposed. For example, Barber et al. (2000, 2002) found a large genetic break between samples of shrimp (*Haptosquilla pulchella*) from north and south of the Flores and Java Seas and suggested that the genetic break resulted from restricted gene flow between regions during lowered sea levels of the Pleistocene. Studies such as this show that a variety of differentiation patterns exists among marine invertebrates: some species have little or no genetic structure (e.g., *Holothuria nobilis*, Uthicke and Benzie 2003; *Centrostephanus rodgersii*, Banks et al. 2007; *Mactromeris polynyma*, Cassista and Hart 2007), while others have pronounced genetic structure (e.g., *Penaeus monodon*, Duda and Palumbi 1999; *Cellana ornata*, Goldstien et al. 2006b; *Haliotis asinina*, Imron et al. 2007).

A key component of gene flow is dispersal. For primarily sedentary marine invertebrates with indirect development, the majority of dispersal occurs during the larval stage. Larvae were once thought to mimic passive particles, and populations were thought to be “open” with extensive larval exchange (reviewed in Levin 2006); however, finding genetic structure among oceanic fauna suggests that the ocean is not a ubiquitous medium in which simply the presence of pelagic larvae results in high levels of gene flow, thereby maintaining genetic connectivity and obliterating the effects of genetic drift and local adaptation (Palumbi 1992; Levin 2006). Instead, a combination of abiotic and biotic factors can limit dispersal and lead to the structuring of marine populations (Palumbi 1992, 1994; Féral 2002; Sponaugle et al. 2002).

Without any obvious barriers to gene flow, very large distances simulate allopatric conditions for diversification. For instance, populations of the echinoderm *Strongylocentrotus droebachiensis* are genetically homogenous on the scale of hundreds of kilometers, but populations separated on the order of thousands of kilometers show significant levels of genetic divergence (Palumbi and Wilson 1990; Addison and Hart 2004). Here, the very large distances act as physical barrier to gene flow. Geologic features also create physical barriers

that significantly limit gene flow between populations. For example, divergence patterns of deep-sea mussels (*Bathymodiolus* spp.) correspond to the topography of the ocean floor (Won et al. 2003). Interlinked with geologic features are hydrographic features, i.e., currents. Hydrographic features can manipulate the spatial and temporal patterns of dispersal and, therefore, gene flow (Davis and Butler 1989; Gaines and Bertness 1992; Gilg and Hilbish 2003). Off the California coast, the strong, southward flowing Californian Current supposedly prevents northward migration of pelagic larvae. Species with pelagic larval stages (e.g., barnacles, *Balanus glandula*, and sea urchins, *Strongylocentrotus purpuratus*) are subjected to the current and have southward patterns of gene flow, whereas gene flow for species with crawling larval stages (e.g., the gastropod, *Nucella emarginata*) is independent of current direction (Wares et al. 2001). In the above examples, the geologic and hydrographic features are contemporary barriers; however as was the case with *H. pulcella* (Barber et al. 2002), patterns of genetic differentiation evident in contemporary populations might be remnants of past geologic or hydrographic features (Benzie 1999).

In the absence of extreme distance or apparent contemporary barriers to gene flow, population differentiation can still occur. Despite being evolutionarily constrained to develop indirectly, organisms are not constrained to disperse, and selection acts on dispersal ability (Hedgecock 1986). Intrinsic dispersal ability depends on characteristics such as larval feeding strategies, adult brooding behavior, and larval behavior. Among other effects, larval feeding strategies influence the length of time pelagic stages are free to disperse and, hence, the amount of genetic structuring. In general, lecithotrophic larvae, which feed off a yolk supply within the egg, have shorter pelagic larval durations than planktotrophic larvae, which feed on other plankton from the water column. Comparisons of genetic differentiation between species with lecithotropic larvae and species with planktotrophic larvae indicate more genetic structure in species with lecithotrophic larvae (e.g., *Adalaria proxima* vs. *Goniodoris nodosa*, Todd et al. 1998, Todd 1998; *Celleporella hyalina* vs. *Electra pilosa*, Goldson et al. 2001). Adult brooding behavior, another characteristic that limits dispersal ability, tends to keep larvae near their birthplace and is associated with larger amounts of genetic structure than broadcasting behavior (e.g., *Balanaphyllia elegans* vs. *Paracyathus stearnsii*, Hellberg 1996; *Epiactis lisbethae* vs. *Anthropleura elagantissima*, Edmands and Potts 1997).

Larval behavior, such as locomotion and habitat choice, influences which currents larvae are exposed to, and when and where larvae metamorphose. Larvae have different modes of locomotion (i.e., swimming or crawling). Currents might be less likely to affect the dispersal patterns and genetic connectivity of crawling larvae as was suggested for the

California current and *N. emarginata* mentioned above (Wares et al. 2001). Larvae's ability to swim, even at speeds lower than the local current's speed, can also shape dispersal patterns. For instance, swimming larvae can position themselves in the water column (i.e., swimming against upwelling and downwelling, Genin et al. 2005) changing their exposure to different currents and, therefore, their dispersal path (e.g., *Tellina* spp. and *Mulinia lateralis*, Shanks and Brink 2005). In addition, larvae can differ in habitat choice affecting the genetic composition of potential recruits. Larvae respond to a broad range of stimuli (i.e., sound, light, chemicals, motion, magnetism, and pressure, Kingsford et al. 2002). Depending on habitat, differential settlement from homogenous gene pools can occur and genetically divide populations (e.g., *Mytilus* spp., Dobretsov and Wahl 2001; Bierne et al. 2003).

The above examples indicate how gene flow can be limited via limiting dispersal. However once larvae disperse, the magnitude of the gene flow is not realized until migrants settle, survive, and reproduce in the new population. This will depend on natural selection and variance in reproductive success. Local environment and pathogens select the genotypes that recruit into a population. For example, the distributions of *Mytilus edulis*, *M. galloprovincialis*, and hybrid genotypes correspond to specific tidal zones (Gilg and Hilbish 2000) and the prevalence of a trematode parasite (Coustau et al. 1991). *M. galloprovincialis*-like mussels have stronger byssal attachment, possibly enabling them to cope with high energy environments of the upper tidal zone (Gardner and Skibinski 1991; Willis and Skibinski 1992). After settlement of a homogenous cohort, *M. galloprovincialis* alleles are selected to recruit higher in the tidal zone (Wilhelm and Hilbish 1998; Gilg and Hilbish 2000; although see Hilbish et al. 2002). Similarly, *Mytilus* spp. genotypes differ in their susceptibility to a trematode parasite. *M. edulis*-like genotypes are more susceptible to infection and, as a result, recruit poorly into areas with high parasite prevalence (Coustau et al. 1991). Individuals that do not survive in the new habitat will not contribute to the future gene pool of the new population, and therefore, the impact of the original gene flow would be slight.

High fecundity and stochasticity of larval survival can lead to extreme variance in the reproductive success of marine organisms (Hedgecock 1994). Pure chance could prevent immigrants from reproducing in a population or larvae from reaching and surviving in a new population. Reproductive success could also vary due to interactions at the level of the sperm and the egg (Palumbi 1994; Swanson and Vacquier 2002). For broadcast spawners, like abalone and sea urchins, gamete recognition proteins play a critical role in successful fertilization and potentially speciation (Palumbi 1994; Vacquier et al. 1995; Swanson and

Vacquier 2002). For such species, differentiation at gamete recognition proteins might prevent the establishment of immigrant genes within a population and reduce the magnitude of gene flow.

As a whole, gene flow for marine invertebrates with a larval stage is quite complex and incorporates periods of dispersal, settlement, and recruitment. Difficulty in tagging individual larvae (for a review of tagging methods see Thorrold et al. 2002 and Levin 2006) has led researchers to rely on inferring barriers to gene flow based on genetic structure. Unless sampling over multiple known cohorts (Thorrold et al. 2002), genetics most likely will only provide a snapshot of population structure that has accumulated over multiple generations. Given the myriad of factors influencing genetic structure, inferences about gene flow should include contemporary and historical knowledge of the biotic and abiotic environment as well as information about the organism of interest.

ABALONE

Haliotis is a diverse genus with 56–70 different species (Lindberg 1992; Geiger 2000). As a group, *Haliotis* has a cosmopolitan distribution (Figure 1.1, Lindberg 1992; Geiger 2000); however, individual species show a much more restricted range (detailed in Geiger 2000). Haliotids (or abalone) are invertebrate macrograzers that inhabit shallow, subtidal, tropical and temperate rocky reefs (Geiger 2000). They are characterized by flat shells with a row of respiratory pores called tremata, paired comb-like gills, and well-developed epipodium (Lindberg 1992). Abalone species vary according to shell, trematal, and epipodial morphology (summarized in Table 4.1 p35 of Tissot 1992). Morphologies are most likely related to habitat (Tissot 1992). For instance, shell and trematal structure are probably influenced by water movement: smoother shells experience less drag force and are associated with shallower and more exposed habitats than more structured shells (Tissot 1992). A combination of heritability and selective breeding projects are attempting to distinguish the amounts of environmental and genetic components in determining abalone morphology and physiology (Jonasson et al. 1999; Lucas et al. 2006; Hara and Sekino 2007; Kube et al. 2007).

In general, adult abalone are long-lived (e.g., H. *corrugata* live for at least 16 years, Gluyas-Millan and Talavera-Maya 2003; *H. fulgens* live for at least 20 years, Gluyas-Millan and Talavera-Maya 2003; *H. rufescens* live for at least 30 years, Rogers-Bennett et al. 2007) and have low natural mortality (e.g., 0.11–0.23 per year for *H. rufescens*, Rogers-Bennett et al. 2007; 0.29–0.36 per year for *H. kamtschatkana*, Zhang et al. 2007). Growth depends on a variety of factors including food supply (Day and Fleming 1992), diet (Serviere-Zaragoza et

al. 2001), and water temperature (Steinarsson and Imsland 2003). Adult abalone are considered sedentary; although, they do move particularly in response to habitat quality and lack of food (e.g., *H. rubra*, Dixon et al. 1998; *H. midae*, Tarr 1995; *H. tuberculata* and *H. discus hannai*, Werner et al. 1995). They are dioecious broadcast spawners found at equal numbers of males and females in the wild (e.g., *H. midae*, Fielding 1995; *H. asinina*, Capinpin et al. 1998; *H. laevigata*, Wells and Mulvay 1995). Some species aggregate (e.g., *H. kamtschatkana*, Breen and Adkins 1980; *H. laevigata*, Shepherd 1986; *H. rubra*, Officer et al. 2001; *H. sorenseni*, Hobday et al. 2001), which could help maintain the high densities of sperm and eggs needed for spawning success (Clavier 1992; Babcock and Keesing 1999).

Abalone are mature at species-specific sizes that are dependent on sex and environment and often require several years of growth to obtain (for examples see Shepherd et al. 1995; Wells and Mulvay 1995; Capinpin et al. 1998; Campbell et al. 2003; Rogers-Bennett et al. 2004). Spawning occurs annually with the periodicity, duration, and synchrony varying within and between species (for examples see McShane 1992 and references therein; Wells and Mulvay 1995; Wood and Buxton 1996; Counihan et al. 2001; Onitsuka et al. 2007). Fertilization is species-specific (Vacquier et al. 1990; Lyon and Vacquier 1999); however, hybrids occur in nature at small frequencies (Brown 1995) and can be produced artificially (Hoshikawa et al. 1998; Ibarra et al. 2005; Ahmed et al. 2008). Abalone have high fecundity. For example, maximum productivity of an adult *H. rufrescens* is 2,850,000 eggs/year (Rogers-Bennett et al. 2004). Fecundity increases with size (Sainsbury 1982a; Worthington and Andrew 1997; Rogers-Bennett et al. 2004); although, Rogers-Bennett (2004) reported that fecundity in *H. rufescens* decreased with shell lengths larger than 215 mm.

The limited movement of adults implies that dispersal occurs primarily through pelagic gametes and larvae. Often within 24 hours of fertilization, the eggs hatch to produce upward swimming trochophore larvae, which subsequently become downward swimming veliger larvae (reviewed in McShane 1992). Abalone larvae are lecithotrophic (although see Jaeckle and Manahan 1992; Shilling et al. 1996) and exist in the water column for 3–10 days (McShane 1992; although *H. iris* can suspend metamorphosis for up to 34 days Roberts and Lapworth 2001). Their spherical shape and cilia-based locomotion suggest that abalone larvae are passively dispersed in water columns with vertical mixing (McShane 1992). Abalone larval dispersal would, therefore, depend on the local hydrodynamic environment. Both McShane and Smith (1988) and Shepherd et al. (Shepherd et al. 1992) have suggested the retention of larvae due to eddies as explanations for patterns of abalone recruitment.

Larvae require an exogenous cue for settling (Morse 1990). In cultured abalone, cues include crustose coralline red algae, various microalgae and bacteria, and abalone mucus; however, crustose red algae appear to be the only effective cue in natural environments (Morse 1992 and references therein; reviewed in Roberts 2001). Crustose red algae produce a molecule that mimics gamma-aminobutyric acid (GABA), a neurotransmitter. Abalone have external receptors that bind to the GABA mimic promoting settlement and metamorphosis, and the receptivity of these receptors is regulated by lysine and related diamino acids (reviewed in Morse 1992). Survival of larval and post settlement stages are probably low due to predation (Naylor and McShane 1997), starvation (Sasaki and Shepherd 2001; Roberts et al. 2004), and hydrology (Naylor and McShane 2001). Recruitment is variable and depends on the numbers of abalone settling, as well as competition, predation, and environmental disturbance (McShane 1992).

Systematics

Abalone belong to the molluscan class Gastropoda and the superorder Archaeogastropoda, which is sister to all other gastropods (Winnepenninckx et al. 1998). Within archaeogastropoda, abalone species are grouped into the suborder Vetigastropoda (which also includes top-snails, keyhole limpets, and turban shells) and then into the monophyletic family Haliotidae (Geiger and Thacker 2005). *Haliotis* species are the only members of Haliotidae. The relationships between Haliotidae and other Vetigastropods are unresolved; however, Haliotidae tends to group with Fissurellidae (key hole limpets), Lepetodrilidae, Scissurellidae (little slit shells), Turbinidae (turban snails), Trochidae (top snails), and Skenidae (Geiger and Thacker 2005; Williams and Ozawa 2006).

Within the genus *Haliotis*, a number of evolutionary relationships have been hypothesized based on a variety of approaches. Geiger's (1999) examination of morphological characters indicated that many characters were plastic and not useful for constructing a phylogeny. As a result, most investigations of abalone phylogenetics have relied on molecules to infer relationships (e.g., Lee and Vacquier 1995; Geiger 2000; Coleman and Vacquier 2002; Estes et al. 2005; Degnan et al. 2006; Streit et al. 2006). Regardless of the data used, abalone species tended to group according to broad geographic locations with some exceptions (Figures 1.1 and 1.2). North Pacific and North American abalone consistently formed a clade (Figure 1.2A, Lee and Vacquier 1995; Figure 1.2C, Geiger 2000; Figure 1.2F, Coleman and Vacquier 2002; Figure 1.2D, Estes et al. 2005; Figure 1.2E, Degnan et al. 2006; Figure 1.2 G, Streit et al. 2006). Endemic Australian abalone also

formed a clade (Figure 1.2A, Lee and Vacquier 1995; Figure 1.2C, Geiger 2000; Figure 1.2F, Coleman and Vacquier 2002; Figure 1.2D, Estes et al. 2005; Figure 1.2E, Degnan et al. 2006).

In contrast, Indo-Pacific species, distributed from Australia to Japan and as far west as Africa (Figure 1.1), appeared to be polyphyletic (Figure 1.2A, Lee and Vacquier 1995; Figure 1.2D, Estes et al. 2005). Degnan et al.'s (2006) analysis, which concentrated on Indo-Pacific species, indicated that cryptic speciation and hybridization might be occurring in Indo-Pacific abalone and that most of the Indo-Pacific abalone were distinct from the endemic Australian species (Figure 1.2E). The Australian and Indo-Pacific species also tended to group with the Mediterranean *H. tuberculata* (Figure 1.2A, Lee and Vacquier 1995; Figure 1.2C, Geiger 2000; Figure 1.2D, Estes et al. 2005).

Southern African and New Zealand species do not have a consistent placement across trees (Figures 1.1 and 1.2). Allozymes (Figure 1.2B) clustered the South African *H. midae* and the New Zealand *H. iris* into a monophyletic group within the endemic Australian clade (Brown and Murray 1992a); otherwise, mitochondrial and nuclear DNA clustered *H. midae* sister to the endemic Australian clade (Figure 1.2A, Lee and Vacquier 1995; Figure 1.2F, Coleman and Vacquier 2002; Figure 1.2D, Estes et al. 2005; Figure 1.2E, Degnan et al. 2006). Placement of *H. iris* remains unresolved. Ribosomal DNA internal transcribed spacer (ITS) sequences indicated *H. iris* is sister to all other haliotids (Figure 1.2F, Coleman and Vacquier 2002). Using 257–318 bp of ITS1 and 289–303 bp of ITS2, Coleman et al. (2002) found that *H. iris* was very distant from all the other species: it was the only species with a transition in a highly conserved region and multiple compensatory base changes. Mitochondrial cytochrome oxidase II data indicated *H. iris* was sister to the Indo-Pacific and Australian abalone (Figure 1.2E, Degnan et al. 2006). Amalgamating data from the mitochondrial 16S and cytochrome oxidase I with nuclear ITS and lysin, Estes et al.'s (2005) analysis grouped *H. iris* with North American and North Pacific species (Figure 1.2D).

Another New Zealand species, *H. australis*, tended to group with the endemic Australian and Indo-Pacific abalone (Figure 1.2C, E, Geiger 2000; Degnan et al. 2006). However like *H. iris*, Estes et al. (2005) analysis grouped *H. australis* with the North American and North Pacific clade (Figure 1.2D). The third New Zealand species, *H. virginea*, either grouped with *H. iris* (Figure 1.2E, Degnan et al. 2006; Wang et al. 2006; Clarke 2001) or with the Indo-Pacific and Australian abalone (Figure 1.2B, Brown and Murray 1992a; Figure 1.2C, Geiger 2000).

Noticeably the divergences between the broad geographic regions are larger and have more support than the divergences within these regions. Furthermore, the species within the

regions (based on Geiger 2000) have overlapping distributions. The deep divergences between regions might reflect ancient isolation rather than increased rates of evolution. Using the formation of the Isthmus of Panama 3–3.5 mya, Coleman and Vacquier (2002) dated the separation of *H. robertii* and *H. pourtalessii* to calculate a rate of molecular evolution (Figure 1.2 F). Applying this rate across the tree, the common ancestor between *H. iris* and the other haliotids dates to ca. 83–111 mya; although, the oldest abalone fossil found in New Zealand dates back to the Miocene (23–5 mya, Hertlein 1937 as cited in Lindberg 1992).

A limited fossil record also supports an ancient origin for the *Haliotis* genus. Unfortunately, scarcity, poor preservation environments (high energy and rocky), argonite shells, and plastic shell morphology plague abalone paleontology (reviewed in Geiger and Groves 1999). Of the 35 fossil species known, most were identified from a single specimen, and whether these specimens represent extant species is unknown. Possible haliotid fossils date back to the Triassic (250–290 mya), but the first certain haliotid fossil was found in Californian sediments dating to >66 mya and is most similar to New Zealand's *H. iris* (Anderson 1902 and Durham 1979 as cited in Geiger 1999). By the Miocene, haliotids were distributed around the world (Lindberg 1992).

Assessing genetic structure

Hey and Machado (2003) partitioned the study of genetic structure into three categories: traditional population genetics, phylogeography, and hybrid approaches. Although these approaches strive to answer questions regarding genetic structure, they are inherently different. Founded in the works of Sewall Wright and Ronald A. Fisher, classical population genetics approaches involve using allele frequency and sequence polymorphism data to calculate summary statistics. The summary statistics rely on specific mathematical models such as an infinite island (Wright 1931; Slatkin 1985; Nagylaki 1998) or a coalescent (Kingman 1982; Hudson 1990; Nordborg 2003). Structure is then assessed by interpreting summary statistics in light of the underlying models (Hey and Machado 2003). In contrast, phylogeographic approaches involve interpreting gene genealogies in term of geographic distributions of individuals (Avise et al. 1987). These approaches are frequently applied to data from which gene trees can be easily estimated, e.g. mitochondrial DNA because it lacks recombination. Unlike summary statistics, gene trees do not depend on specific models and may be interpreted at face value; however, gene tree approaches suffer from large stochastic variance among trees derived from different genes. Hybrid approaches estimate model parameters from gene trees (Hey and Machado 2003).

Studies of abalone genetic structure have employed population genetic or phylogeographic methods. The population genetic assessments using summary statistics have remained classical and have not yet applied coalescent-based models. The review below focuses on the presence and absence of genetic structure determined by whether F_{ST} or Φ_{ST} differed significantly from 0 via randomizations. These values measure genetic differences among subpopulations relative to the total population (Hartl and Clark 1997). Depending on the nature of the data, F_{ST} values are calculated in different ways (Excoffier 2003). As a result, studies report F_{ST} analogues, but the analogues essentially measure the same concept. For simplicity, the analogues presented in various papers are reported here as F_{ST} for allelic data and Φ_{ST} for sequence data. F_{ST} and Φ_{ST} values are useful but should be interpreted cautiously (see Whitlock and McCauley 1999; Neigel 2002; Pearse and Crandall 2004). F_{ST} and Φ_{ST} provide a measure of the extent of population subdivision, but are not easily translated into measures of geneflow (Whitlock and McCauley 1999), cannot differentiate between processes, i.e. contemporary or prehistorical barriers to gene flow (Pearse and Crandall 2004), and are biased by demographic history (Whitlock and McCauley 1999). Neither the magnitudes of F_{ST} nor Φ_{ST} are readily comparable because they depend on characteristics of the markers used (i.e., mutation rate and polymorphism, Hedrick 1999). In instances when F_{ST} or Φ_{ST} values are not given, structure is assessed according to the original authors' interpretation.

Genetic structure of abalone

Given their commercial importance, *Haliotis* spp. are popular mollusks for exploring genetic connectivity. Withler (2000) pointed out the paucity of population genetic data for abalone, and researchers, hoping to elucidate population genetic structure, have since responded with a surge of articles employing various molecular markers across different abalone species (Table 1.1 and 1.2). Despite the increasing number of studies examining abalone population genetics, two clear trends among *Haliotis* spp. are evident. First, haliotids harbor lots of intraspecific variation. Second, many studies report either panmixia or weak but significant genetic structure. Otherwise findings and interpretations are rather diverse, and inconsistencies among sympatric or monophyletic species could be due to species-specific differences, sampling scale, and analyses employed.

North America

On the western coast of North America, seven species of abalone have overlapping ranges. Driven by commercial and management interests, economically important abalone species (*H. rufescens*, *H. cracherodii*, *H. fulgens*, *H. corrugata*, and *H. kamtschatkana*) have been studied from a genetic perspective. Of the North American species, *H. rufescens* has attracted the most attention. Interest in *H. rufescens* began with attempts to evaluate a stock enhancement that occurred in 1979 at San Miguel Island, California (Gaffney et al. 1996; Burton and Tegner 2000). Gaffney et al. (1996) compared four allozymes between southern California abalone sampled in 1979 and northern California abalone sampled in 1992 and found no significant differences between the two groups. Burton and Tegner (2000) also evaluated the 'enhanced' southern California population and two northern sites and found no genetic differentiation with three allozyme loci (same loci used by Gaffney et al. 1996) and 484 bp of mitochondrial cytochrome oxidase I (mtCOI).

Gaffney et al. (1996) and Burton and Tegner (2000) focused on answering questions regarding stock enhancement, and as result, their studies were limited in the use of molecular markers and in sampling scheme. Further research on *H. rufescens* has attempted to apply a variety of markers and broaden the geographic scope. Kirby et al. (1998) screened 39 northern and 35 southern abalone with one microsatellite locus (28 alleles) and noted five private alleles in the northern group and one private allele in the southern group but did not further explore this potential split. Gruenthal et al. (2007) has conducted the most extensive assessment of population genetic structure in *H. rufescens*. Sampling nine localities spanning about 1300 km of coastline, they examined mtCOI and polymorphism at five microsatellite loci. Like Burton and Tegner (2000), Gruenthal et al. (2007) found no differentiation among sampling localities based on 483 bp of mtCOI, while the five microsatellites indicated very weak population structure (global $F_{ST} = 0.002$; $p = 0.002$). Additional assessment of a subset of five sampling sites with 163 polymorphic AFLP markers further rejected homogeneity among sampling sites, but no obvious pattern was evident (Gruenthal et al. 2007).

Compared to *H. rufescens*, black abalone (*H. cracherodii*) exhibited more genetic structure as estimated with allozymes ($F_{ST} = 0.039$, $p < 0.001$, Hamm and Burton 2000). Allele frequencies at three loci were heterogenous across seven sampling locations spanning over 300 km; however, no clear pattern to the heterogeneity existed, nor was structure detected in 382 bp of mtCOI (Hamm and Burton 2000). Chambers et al. (2006) extended sampling of *H. cracherodii* to include four islands offshore of southern California. Based on five allozymes, sampling sites had weak but significant genetic differentiation ($F_{ST} = 0.008$; $p < 0.001$). Hierarchical tests of genetic variation grouped sampling localities into offshore,

nearshore, and mainland sites. Gruenthal and Burton (2008) screened Hamm and Burton's (2000) and Chambers et al.'s (2006) samples (spanning around 750 km) with mtCOI, four microsatellites, and 142 AFLPs. Microsatellites indicated a panmictic population, while mtCOI (global $\Phi_{ST} = 0.014$, $p = 0.010$) and AFLPs (global $F_{ST} = 0.044$, $p = 0.001$) indicated slight but significant structure. Gruenthal and Burton (2008) did not find any pattern of genetic structure based on mtCOI but did find a slight pattern of isolation by distance based on AFLPs (Gruenthal and Burton 2008).

Population genetic studies of *H. fulgens* and *H. corrugata* are limited to the coasts of Baja California. Over a small geographic scale (84 km), Zúñiga et al. (2000) concluded that five sampling localities of *H. fulgens* were homogenous based on seven allozymes ($F_{ST} = 0.0461$; although their conclusion was based on $0.05 > p > 0.01$). Gutiérrez-Gonzalez (2000) reported an even lower F_{ST} (0.022) based on 11 polymorphic allozymes for four sampling locales more than 450 km apart. Sampling over eight locations spanning the same distance as Gutiérrez-Gonzalez (2000) and an additional sample from Isla Guadalupe (334 km from the nearest mainland site), Guitiérrez-Gonzalez et al. (2007) found significant but weak genetic structure using four microsatellite loci ($F_{ST} = 0.00062$, $p = 0.002$). This significant structuring was due to inclusion of samples from Isla Guadalupe; removal of the Isla Guadalupe samples resulted in no significant structure of mainland sites and, thus, was consistent with the pattern observed in their earlier study using allozymes (Gutiérrez-Gonzalez 2000). Research on *H. corrugata* has been limited to Cedros and San Benitos Islands just off the Baja coast (del Río Portilla 2000; del Río Portilla and González-Avilés 2001). Based on eight allozymes, six sampling sites were considered distinct with an overall ($F_{ST} = 0.093$, $p < 0.05$) and 12 out of 15 significant pairwise comparisons (del Río Portilla and González-Avilés 2001).

The final North American species, *H. kamtschatkana*, has a distribution stretching from Baja California to Alaska; however, only northern abalone have been assessed for population genetic structure. Sampling 31 sites from Queen Charlotte Island to Vancouver Island and one site from Alaska, Withler et al. (2003) found significant genetic structure using eight microsatellite loci ($F_{ST} = 0.002$, $p < 0.05$). Hierarchical analysis indicated that 99.6% of the variation occurred within samples and 0.4% among samples. Of the among sample variance, half was attributed to differences between Queen Charlotte Island and Alaskan samples and coastal British Colombia samples. Isolation by distance analyses were significant with the inclusion of the Queen Charlotte Island and Alaskan samples.

In general, the presence of structure among sampling sites in the above studies was variable, and when structure was detected, the F_{ST} values were low. Studies incorporating

hierarchical analyses indicated that much of the variation occurred within samples rather than among samples (e.g., Withler et al. 2003; Chambers et al. 2006; Gruenthal et al. 2007; Gruenthal and Burton 2008). Patterns of structure among samples have been attributed to isolation by distance (e.g., *H. kamtschatkana*, Withler et al. 2003; *H. fulgens*, Gutiérrez-Gonzalez et al. 2007) and regional hydrography (e.g., *H. cracherodii*, Chambers et al. 2006; Hamm and Burton 2000; *H. rufescens*, Gruenthal et al. 2007; *H. fulgens*, Zúñiga et al. 2000; Gutiérrez-Gonzalez et al. 2007; *H. corrugata*, Daugherty et al. 1993). Withler et al. (2003) was the only study to propose a prehistoric explanation for genetic structure. They suggested, as an alternative to contemporary hydrography, that the differences among *H. kamtschatkana* samples might be related to northern marine refuges during the last glaciation.

Studies of *H. cracherodii* and *H. rufescens* had overlapping sampling ranges and provided an opportunity to decipher the species and marker-specific components of any population structure versus the overlying oceanographic and geologic processes that promote differentiation common to both species. According to allozyme data, the *H. cracherodii* population was structured (Hamm and Burton 2000), while the *H. rufescens* was not (Burton and Tegner 2000). Hamm and Burton (2000) noted that this could result from variation in spawning time: larvae of *H. cracherodii*, a seasonal spawner, disperse during the limited oceanographic conditions of the summer, while larvae of *H. rufescens*, a year-round spawner, disperse during a variety of oceanographic conditions. However according to microsatellites and AFLPs, both *H. cracherodii* and *H. rufescens* have very weak but significant genetic structure (Gruenthal and Burton 2008; Gruenthal et al. 2007, respectively).

North Pacific

Five species are limited in their distribution to the northwest Pacific Ocean. Population genetic studies within this region focus on the commercially important *H. discus*. Two subspecies of *H. discus* are recognized: *H. discus hannai* inhabit cold northern waters and *H. discus discus* inhabit warm southern waters (Hara and Sekino 2005). Although they were questioning genetic variation in hatchery strains, Li et al. (2004) included two wild samples from the northeastern coast of Japan in their analysis using six microsatellite markers. No genetic differentiation was evident between the wild samples ($F_{ST} = 0.004$). Sekino et al. (2005) also examined two wild populations of *H. d. hannai* with nine microsatellites. In contrast to Li et al. (2004), Sekino et al. (2005) found significant genetic heterogeneity between sites ($F_{ST} = 0.048$, $p = 0.00$) and accredited the heterogeneity to either isolation by distance or effects of stocking areas with hatchery abalone (discussed below). Hara and Sekino (2005) examined only wild populations of both *H. d. hannai* and *H. d. discus*. As

expected, eight microsatellite markers revealed a significant difference between *H. d. hannai* samples and *H. d. discus* samples (F_{ST} = 0.025, p = 0.000). Pairwise comparisons between sampling locations within each subspecies indicated the same pattern of structuring: the sampling site in the Sea of Japan differed significantly from sites in the Pacific Ocean. Unlike previous allozyme studies, no differences were found among the Pacific sites within each subspecies (Hara and Sekino 2005 and references therein). Hierarchical analyses of variance confirmed this relationship, and also revealed that most of the variance was within populations (approximately 96%, Hara and Sekino 2005).

Indo-Pacific

At least 17 abalone species inhabit Indo-Pacific waters (Geiger 2000). Despite the large number of species, very little intraspecific genetic research has occurred in this region. *H. diversicolor* are distributed from Japan to Australia, but only intraspecfic comparisons have occurred between five sites in Taiwan (Jiang et al. 1995). With two restriction enzymes, Jiang et al. (1995) found three distinct mitochondrial RFLP (Restriction Fragment Length Polymorphisms) patterns, but these RFLP patterns did not correspond clearly with geography. The common RFLP pattern was found in one northeastern and two eastern locations, while the less common RFLP pattern was found in one northeastern and one eastern location.

Intraspecific genetic studies have also been conducted for the tropical abalone *H. varia* and *H. ovina* in Thailand (Klinbunga et al. 2003). In Thailand, *H. varia* is only found in the Andaman Sea on the western coast, and Klinbunga et al. (2003) found no evidence of genetic structure in *H. varia* sampled from two locations on this coast using mitochondrial 16S rDNA RFLPs or nuclear 18S rDNA RFLPs. Unlike *H. varia*, *H. asinina* and *H. ovina* are found on Thailand's eastern (Gulf of Thailand) and western coasts. *H. asinina* have a panmictic population structure: all pairwise F_{ST} based on 16S rDNA RFLPs and 18S rDNA RFLPs were not significant between the coasts. Pairwise comparisons of *H. ovina* tended to support genetic differentiation between the northeastern and western populations. Of the four northeastern vs. western comparisons for *H. ovina*, all were significant using 16S rDNA RFLPs (F_{ST} = 0.9444–1.0000, all $p < 0.0001$, but only one comparison was significant (F_{ST} = 0.1296, p = 0.0035) using 18S rDNA RFLPs (Klinbunga et al. 2003). Instead of panmixia, 113 RAPD (Randomly Amplified Polymorphic DNA) fragments and three microsatellites revealed that *H. asinina* samples from two sites in the Gulf of Thailand were different from a sample on the west coast of Thailand in Andaman Sea (Tang et al. 2005).

H. ovina and *H. asinina* are not limited to Thailand: their distributions span from Japan to Australia (Geiger 2000). Imron et al. (2007) sampled *H. asinina* from localities

throughout the Indo-Pacific and found strong genetic structuring among eastern, western, and Indo-Malay regions. Employing a phylogeographical approach using 482 bp of mtCOII, Imron et al. (2007) showed that differentiation among regions explained 73% of the observed genetic variation. They concluded that the split between eastern Australia and the rest of samples was probably due to a historical allopatric event related to late Pleistocene glaciations (Imron et al. 2007). Within the Indo-Malay region, no population genetic structure was found. Even though this region was split in the Pleistocene, Imron et al. (2007) proposed gene flow following the currents and coral reef habitats has occurred, since sea levels rose, and has obscured the patterns of genetic structure.

Southern Africa

Five species inhabit the southern coast of Africa; two of these are only found east of Cape Agulhas (excluding a specimen from the west coast for *H. speciosa*, Geiger, 2000). Evans et al. (2004b) conducted a population genetic study of *H. midae* distributed east and west of Cape Agulhas. They found weak structure using seven allozymes and differentiation between eastern and western coasts using mitochondrial and microsatellite data, but the exact location of the split could not be pinpointed. Within the western and eastern regions, samples were homogenous.

Australia

The majority of population genetic research on Australian abalone has been exploratory and has concentrated on *H. rubra* (blacklip abalone). *H. rubra* is mostly located along the southern coast of Australia, and its distribution overlaps with seven other endemic abalone and five Indo-Pacific abalone (Geiger 2000). Based on 12 allozymes, Brown (1991) found significant genetic structure among 17 localities along the southern coast of Australia and Tasmania. Allele frequencies differed between neighboring samples and suggested an overall isolation by distance structure (Brown 1991; Brown and Murray 1992b). Brown (1991) noted that the large divergence of isolated populations was potentially due to local recruitment. Temby et al.'s (2007) assessment of 18 sampling sites using three microsatellites also suggested local recruitment as a crucial process explaining small-scale heterogeneity. Unlike other *H. rubra* studies, Temby et al. (2007) sampled locations separated by much smaller distances and found significant differences between samples separated by 100 – 200 m but not between samples separated by tens of kilometers.

Over larger scales, RAPDs, microsatellites, and mitochondrial RFLPs have identified genetic structure among *H. rubra* samples. The limited distribution of rare alleles for two minisatellites hinted at population subdivision for ten *H. rubra* samples along the southern

coast of Victoria (Huang et al. 1997). Huang et al. (2000) followed up this preliminary finding by screening the ten localities with 84 RAPD loci and three microsatellites. Both RAPDs and the microsatellites revealed significant genetic structure (F_{ST} = 0.074, $p < 0.001$ and 0.067, $p < 0.001$, respectively). No genetic structure was found using the minisatellites (F_{ST} = 0.001). However when samples were partitioned into three management zones, both minisatellites and RAPDs indicated differentiation among zones but microsatellites did not. Like Brown's (1991 and 1992) allozyme studies, the RAPDs and microsatellites supported an isolation by distance structure (Huang et al. 2000). Conod (2002) sampled *H. rubra* at four localities around Tasmania and one locality from Victoria to test hypotheses of isolation by distance and the role of the Bass Strait as a barrier to gene flow using RFLP of mitochondrial ND3/COIII and five microsatellites. Mitochondrial RFLP and the microsatellite markers showed no differentiation among the Tasmanian sites, but showed significant differentiation between Tasmanian and Victorian samples (Conod et al. 2002).

In addition to *H. rubra*, *H. laevigata* and *H. roei* have also been the subject of intraspecific genetic studies. Brown and Murray (1992b) sampled *H. laevigata* from eight locations over a much smaller range than the *H. rubra* samples (Brown and Murray 1992b) and screened them for variation at 15 allozyme loci. Like *H. rubra*, *H. laevigata* samples had local heterogeneity with an overall isolation by distance structure. Sampling nine sites from west Australia and one from south Australia, Hancock (2000) applied eight allozyme and found genetic subdivision (F_{ST} = 0.0087, $p < 0.001$) in *H. roei.* Hancock (2000) interpreted the data to mean high levels of gene flow, high levels of local heterogeneity, and an overall isolation by distance structure.

New Zealand

Three endemic species of abalone are found around New Zealand's coast (Geiger 2000). Based on allozymes, *H. iris* appear to be panmictic (Dollimore 1977; Frusin 1982). Dollimore (1977) sampled five locations from the North and South Islands and found no obvious variation in allele frequencies for two allozymes. Frusin's (1982) analysis of structure included a North Island, a South Island, and a Chatham Island location and found no significant variation in the common allele frequency among locations. Smith and McVeigh (2006) conducted a preliminary assessment of population genetic structure of *H. iris* based on a mitochondrial region spanning ATPase8–ATPase6. They found significant haplotype heterogeneity among three mainland locations and a Chatham Island location, but not among mainland locations. Screening the samples with six microsatellite markers indicated structure among all sites with each site being significantly different (Smith and McVeagh 2006).

Genetic structure has also been studied in *H. virginea* (Clarke 2001). *H. virginea* is composed of four subspecies that vary in size and distribution (Lindberg 1992). Sequence variation from mitochondrial 16S revealed no differentiation among North and South Island subspecies and arguable differentiation among these mainland subspecies, the Chatham Islands subspecies, and the Campbell Island subspecies.

Hatchery

Considering their commercial importance for both aquaculture and fishing, population genetic studies on abalone include diversity assessments of hatchery strains. Like other shellfish species (e.g., *Crassostrea gigas*, Hedgecock and Sly 1990; *Tridacna gigas*, Benzie and Williams 1996; *Argopecten irradians*, Blake et al. 1997, Wang et al. 2007; *Coelomactra antiquata*, Kong and Lee 2007), hatchery strains of abalone show reduced genetic variability compared to wild stocks and genetic differentiation from wild stocks (e.g., *H. iris*, Smith and Conroy 1992; *H. midae* and *H. rubra*, Evans et al. 2004a; *H. discus hannai*, Li et al. 2004, Li et al. 2007a). They also show slightly reduced variability from parent to offspring (e.g., *H. fulgens*, Gutiérrez-Gonzalez et al. 2005). Since releasing hatchery bred abalone is common practice, such abalone could bias studies of wild species particularly in areas of known stock enhancement (Sekino et al. 2005; Hara and Sekino 2005). Sekino et al. (2005) claimed that stock enhancement would result in the individuals within samples being more closely related than between samples and called this result the "long term effect of stocking populations." On the other hand, stock enhancement might not be successful and contribute little to current genetic structure (e.g., *H. rufescens*, Burton and Tegner 2000; *H. fulgens*, Gutiérrez-Gonzalez et al. 2005).

Summary

As with the California abalone, most intraspecific studies of other abalone species found weak but significant genetic differentiation, high levels of diversity using markers other than allozymes, and predominantly local heterogeneity without a large-scale pattern. Local heterogeneity has been attributed to local retention of larvae (e.g., *H. rubra*, Brown 1991; *H. roei*, Hancock 2000). Beyond local heterogeneity, significant F_{ST} values have been deciphered as limited gene flow due to geographic distance (e.g., mainland *H. rubra*, Brown 1991; although not for Tasmanian *H. rubra*, Conod et al. 2002), geological features (e.g., *H. discus*, Hara and Sekino 2005), oceanographic currents (e.g., *H. asinina*, Imron et al. 2007) and prehistorical conditions (e.g., *H. asinina*, Imron et al. 2007).

If geological features, oceanographic currents, and prehistoric conditions are barriers to gene flow, then they would be expected to leave similar signatures on other species in the region; however, not many authors go beyond identifying structure. Evans et al. (2004b) and Imron et al. (2007) found considerable levels of differentiation in *H. midae* and *H. asinina*, respectively. *H. midae* appeared to have a genetic split around Cape Agulhas. As the meeting point for the Indian and Atlantic oceans, Cape Agulhas has a complex hydrology (e.g., Largier et al. 1992; Boebel et al. 2003) that would be expected to influence genetic structuring of populations. Unfortunately, Cape Agulhas has not been thoroughly explored for its role in genetic structuring of populations. Unlike Cape Aghulas, genetic structure has been examined in several marine invertebrates throughout the Indo-Pacific (e.g., *Holothuria nobilis*, Uthicke and Benzie 2003; *Echinolittorina* spp., Reid et al. 2006; *Nerita albicilla* and *Nerita plicata*, Crandall et al. 2008). The genetic discontinuity in *H. asinina* (Imron et al. 2007) between the Indian and Pacific basins is similar to splits identified *Penaeus monodon* (Benzie et al. 2002) and *Echinolittorina* spp. (Reid et al. 2006). This concordance lends support for the plio-pleistocene glaciations influencing the genetic structure of marine invertebrates in the Indo-Pacific region.

NEW ZEALAND ABALONE

New Zealand has three endemic abalone species, *H. iris*, *H. australis*, and *H. virginea*. Prized for their colorful shell and tasty meat, New Zealand abalone are commercially and culturally very important. New Zealand abalone (predominantly black-foot paua, *H. iris*) have been fished at close to or over 1000 t from 1987–2006 (FAO 2000). Out of the thirteen countries reporting commercial catch in 2006, New Zealand was the third largest harvester of abalone at 952 t (FAO 2000) with exports reaching $50.0 million (Figure 1.3a, http://www.fish.govt.nz 2006). World catch as a whole from 1970 to 2006 has declined (Figure 1.3b). Remarkably despite the world's decline, New Zealand's harvest has remained relatively constant since 1981 (Figure 1.3c). This, in part, might be due to a well-established quota driven management system (Leiva and Castilla 2001) and/or plethora of research on *H. iris* biology.

The biology of *H. iris* is very similar to the general abalone biology described above and is repeated here for species-specific clarity. *H. iris* inhabit intertidal and subtidal rocky reefs surrounding New Zealand and offshore islands, but is most abundant in the cooler waters south of Cook Strait (Schiel and Breen 1991). *H. iris* aggregate (Poore 1972b; McShane 1996) and are found at depths up to 15 m with most living between 0.5–7.0 m

(Sainsbury 1982a). Adult *H. iris* have been called sedentary; however, mark and recapture data and observations indicate *H. iris* are mobile and active (Poore 1972a; Sainsbury 1982a; Naylor and McShane 2001). *H. iris* feed on either drifting or attached algae with a preference for red algae (Poore 1972a). Their growth is variable and linked to water turbulence (McShane et al. 1994; McShane and Naylor 1995b), water temperature (Naylor et al. 2006; Searle et al. 2006), season (Allen et al. 2006), and diet (Poore 1972a; Stuart and Brown 1994). Size composition of populations are variable, but tends to be skewed toward larger individuals (Sainsbury 1982a; Schiel and Breen 1991). Once abalone shell length reaches 30 mm, natural instantaneous mortality is constant around 0.1 per year (Sainsbury 1982b; McShane and Naylor 1997).

Size at maturation is variable and negatively correlated with mean monthly sea surface temperature (Naylor et al. 2006) and positively correlated water turbulence (McShane and Naylor 1995b). Fecundity is high (11,253,000 eggs for an abalone shell length of 155mm, Poore 1973). Fecundity increases with length up to about 120 mm and is variable across larger abalone (Sainsbury 1982a). Spawning tends to occur during late summer (Poore 1973); however, spawning events are not always annual and timing can vary by site (Poore 1973; Sainsbury 1982a). For instance, Hooker and Creese (1995) reported three spawning events at a North Island locality: a major spawning event occurred during June–July 1987, another major spawning event occurred during September–October 1987, and a minor spawning event occurred in February–March 1987.

Fertilized eggs are negatively buoyant (Tong et al. 1992). After 12 hours, zygotes develop into swimming larvae (trocophore and veliger stages) for four to eight days, and larvae are competent to metamorphose at five days post-fertilization (Tong et al. 1992). However, laboratory reared *H. iris* are capable of delaying metamorphosis until an age of 34 days at 17 °C (Roberts and Lapworth 2001). Post-settlement survival and growth are high (>80 % and >20 μm/day, respectively) for larvae that metamorphose before or at 22 days post-fertilization (Roberts and Lapworth 2001). Larval survival, settlement, and recruitment are variable and influenced by a variety of abiotic and biotic factors. For instance, larval development and survival is negatively impacted by sediment concentration (Phillips and Shima 2006). Larval settlement is affected by food available: larvae prefer settling on crustose coralline algae (Roberts et al. 2004). Post settlement survival varies according to presence of conspecifics (Naylor and McShane 2001), depth (McShane and Naylor 1995a), predation (Naylor and McShane 1997), wave exposure (Naylor and McShane 2001), and grazing by conspecific adults (Naylor and McShane 2001). The myriad of factors affecting larval

survival and settlement and juvenile survival could result in variable recruitment over time and space, as observed by Sainsbury (1982a)

Aims

For marine invertebrates, population differentiation occurs by a variety extrinsic and intrinsic factors. A comprehensive study to tease apart these factors and their contributions to creating and maintaining genetic structure would require long term ecological and genetic studies that incorporate sampling of larvae, juveniles, and adults over multiple years. Creative sampling strategies, comparative methods, and new sequencing technologies offer novel alternatives. This thesis was initially designed to incorporate such alternative approaches; however, it became limited due to notorious technical difficulties of working with *Haliotis iris* DNA. The result is a study examining the genetic structure of New Zealand's *H. iris* with an emphasis on a potential extrinsic barrier to gene flow, the Cook Strait region, and a potential intrinsic factor, sperm and egg recognition proteins. This study is important because it 1) lays the foundation for future evolutionary and comparative studies with other New Zealand abalone, 2) is the first study of a New Zealand marine invertebrate that intentionally uses a non-neutral genetic marker to study patterns of differentiation, 3) uses markers that are applicable to selective breeding projects, and 4) has applications in the conservation and management of a commercially and culturally important species.

Abalone were originally chosen to examine the factors effecting marine population differentiation because they possess well-characterized gamete recognition proteins enabling the study of intrinsic, in addition to extrinsic, factors of population differentiation. They are also a commercially important species making funding and a relatively large amount of information regarding their life history, behavior, and physiology readily available. Studying abalone in New Zealand was particularly tantalizing because three species occupying slightly different niches coexist around New Zealand offering the potential for a comparative framework. Furthermore, a fair amount of literature regarding New Zealand's oceanography and marine invertebrate genetic structures exists enabling an a priori framework to test population differentiation.

Given the similarity of *H. iris* biology to general abalone biology, *H. iris* could be predicted to have similar population genetic trends, such as large amounts of genetic variability and weak genetic structure that is only detectable with either very polymorphic loci or a large number of loci (e.g., see Gruenthal et al. 2007; Gruenthal and Burton 2008). On the other hand a few abalone studies detected considerable genetic structure related to complex

contemporary or prehistorical oceanographic conditions (e.g., see Evans et al. 2004b; Imron et al. 2007). In the case of *H. asinina* (Imron et al. 2007), the population genetic structure corresponded with genetic structures of other Indo-Pacific marine invertebrates (e.g., *Penaeus monodon*, Benzie et al. 2002; *Echinolittorina* spp., Reid et al. 2006). Similarly, *H. iris* might have a pronounced genetic structure that would be concordant with New Zealand's oceanography and/or the genetic structures of other New Zealand coastal marine invertebrates.

New Zealand's complex marine environment could promote genetic structuring in *H. iris*. Situated between the Tasman Sea and the Pacific and Southern Oceans, New Zealand is surrounded by six major offshore currents (Figure 1.4, summarized in Heath 1985; Hume et al. 1992; Uddstrom and Oien 1999; Laing and Chiswell 2003). In general, the waters surrounding the North Island are warmer and more saline than the waters surrounding the South Island. As the Tasman Front passes New Zealand, it generates the major North Island currents, the East Auckland Current (EAUC) and the West Auckland Current (WAUC), that bathe the North Island in warm subtropical water. As the EAUC passes the East Cape, water either continues east towards the Pacific or flows south along the eastern coast of the North Island creating the slightly cooler and weaker East Cape Current (ECC). Along the west coast of the North Island, the WAUC also flows south and is slightly cooler and has a weaker and more variable flow than EAUC (Heath 1985; Hume et al. 1992; Uddstrom and Oien 1999; Laing and Chiswell 2003).

In contrast to the North Island, the water surrounding the South Island is derived from the cooler, less saline Tasman Sea. The major currents around the South Island are the Westland Current (WC) and the Southland Current (SC). The WC flows north along the west coast, and its flow is weaker and more variable than the SC which begins on the west coast and flows around the south and east coasts until it reaches Banks Peninsula. At Banks Peninsula part of the SC heads east along the Chatham rise, while the remaining water flows north. Finally, the D'Urville Current (DC) flows from west to east through Cook Strait. Water movements through the strait are further complicated with large amounts of tidal mixing, upwelling, eddies, and river plumes (Harris 1990; Vincent et al. 1991). New Zealand's diverse marine environment has stimulated a body of literature examining connectivity of coastal marine invertebrates. A consistent pattern emerging from literature of New Zealand coastal marine invertebrates is that genetic structure exists across the Cook Strait region (Ayers and Waters 2005; Veale 2007; Goldstien et al. 2006b).

The first aim of this thesis is to assess the population genetic structure of *H. iris* sampled from around New Zealand. In contrast to the previous studies by Dollimore (1977), Frusin (1982), and Smith and McVeagh (2006), this study samples more intensely, applies mitochondrial DNA and microsatellites as molecular markers, and uses population genetic and phylogeographic approaches to identify genetic patterns, particularly in reference to the Cook Strait. **Chapter 2** explores mitochondrial variation among *H. iris* samples from around New Zealand and examines the results in comparison to other New Zealand marine invertebrates. The mitochondrial analysis is then followed by a similar study employing nuclear, microsatellite markers. **Chapter 3** records the isolation and characterization of microsatellite markers for *H. iris*. **Chapter 4** applies the microsatellites to study nuclear variation among samples of New Zealand abalone and compares the observed patterns to those obtained for mitochondrial DNA.

The second aim of this thesis is to examine variation in *H. iris* gamete recognition gene lysin. Abalone gamete recognition proteins are well characterized and could contribute to speciation (Panhuis et al. 2006). Unlike other marine invertebrate gamete recognition proteins (e.g., sea urchin bindin, Metz and Palumbi 1996, Palumbi 1999, Geyer and Palumbi 2003; mussel lysin M7, Riginos and McDonald 2003, Riginos et al. 2006, Springer and Crespi 2007; oyster bindin, Moy et al. 2008), lysin shows very little within species variation and is thought to be under intense selection. However, lysin has only been studied in a small number of individuals from different species (e.g., Lee and Vacquier 1995; Metz et al. 1998b; Swanson and Vacquier 1998; Clark et al. 2007). Given the high levels of intraspecific variation in abalone and the potential of population genetic structure in *H. iris*, *H. iris* lysin could be variable and indicate a structure similar to other molecular markers. **Chapter 5** recounts attempts to sequence lysin and characterizes variation and genetic structure using lysin coding and noncoding sequence. Since nuclear sequence data evolves differently from mitochondria and microsatellites, **Chapter 6** describes variation in the neutral nuclear Gα protein intron in order to interpret the intraspecific variation in lysin.

2. Genetic structure across Cook Strait

ABSTRACT

A consistent pattern of genetic structure persists among New Zealand coastal marine invertebrates: specifically, a north-south split occurs in the Cook Strait region. The hydrography of the Cook Strait region is complex. Potentially, upwelling regions off the southeastern and southwestern shores of Cook Strait act as a barrier to gene flow; however, the few studies that date the genetic split indicate that the splits origin is older than the contemporary hydrography. Regardless of the nature of the north-south split, its existence in other coastal marine invertebrates provides an a priori framework for examining abalone, *Haliotis iris*, population genetic structure. The presence of population genetic structure was tested in reference to the Cook Strait region using 459 bp of the mitochondrial cytochrome oxidase I and 596 bp of a region spanning ATPase8–ATPase6 amplified in *H. iris* from 25 locations around New Zealand. Assuming an a priori structure, AMOVAs indicated that the Chatham Islands sample was significantly different from the mainland samples and that the mainland North Island samples were significantly different from the South Island samples. The north-south split is reflected in the abundance of private haplotypes and high haplotype diversity in northern samples. Potentially either increased mutation rates in North Island *H. iris* or fisheries induced bottlenecks in South Island *H. iris* are driving the disparity in haplotype diversity.

INTRODUCTION

New Zealand's isolation, oceanography, and abundant fisheries have sparked a considerable amount of marine invertebrate population genetic research (Table 2.1 and thoroughly reviewed in Goldstien 2005 and Veale 2007). The studies have varied widely in geographic scope, number of individuals sampled, type of marker and number of loci used, and amount and pattern of genetic structure identified. Some species were panmictic (e.g., *Jasus edwardsii*, Smith et al. 1980; Ovenden et al. 1992; *Crassostrea giga*, Smith et al. 1986) while other species had pronounced genetic structures (e.g., *Cellana ornata*, Goldstien 2005, Goldstien et al. 2006b; *Sypharochiton pelliserpentis*, Veale 2007). Patterns of genetic structure have included differentiation between northern and eastern samples of the North Island and other North Island samples (e.g., *Patriella regularis*, Waters and Roy 2004), differentiation between northern and eastern samples of the North Island from remaining North Island and South Island samples (*Paracorophium lucasi* and *P. excavatum*, Schnabel et al. 2000; Stevens and Hogg 2004), and a gradual differentiation by distance (*Actinia tenebrosa*, Veale 2007).

New Zealand's dynamic offshore environment offers a variety of features that could promote differentiation. As described in Chapter 1 (Figure 1.4), New Zealand is surrounded by six major surface currents: East Auckland Current, West Auckland Current, East Cape

Current, Westland Current, Southland Current, and the D'Urville Current (Figure 1.4). These currents differ in flow, temperature, and salinity, and correlate with the genetic structure of limpets (Goldstien 2005), amphipods (Stevens and Hogg 2004), and brittle stars (Sponer and Roy 2002). Currents, along with sea surface temperature, have also been suggested as determinants for the composition of marine provinces for fish (Moreland 1959; Francis 1996), mollusks (Powell 1955), echinoderms (Pawson 1961), and seaweeds (Moore 1961).

Located on the southwest of the South Island are 14 deep water fiords carved by glaciers 14,000 years ago (Smith 2001). Although the fiords were not associated with differentiation in *Patriella regularis* (Waters and Roy 2004; Ayers and Waters 2005) and *Perna canaliculus* (Apte and Gardner 2001; Apte and Gardner 2002; Star et al. 2003), fiordic samples of sea urchins, *Evechinus chloroticus* (Mladenov et al. 1997; Perrin 2002), and sea stars, *Coscinasterias muricata* (Sköld et al. 2003; Perrin et al. 2004), were significantly different from other samples around New Zealand. Both these species also had significant genetic structure within the fiords. *E. chloroticus* sampled from the inner and outer fiords were different (Perrin 2002). For *C. muricata*, no discernable pattern was evident based on allozymes (Sköld et al. 2003); however, analysis of the mtDNA control region indicated two lineages, a zone of mixing, and a partial isolation by distance structure (Perrin et al. 2004).

Additional permanent and ephemeral hydrological features around New Zealand that could also affect population genetic structure include eddies, wind-induced upwelling events, and river plumes (Heath 1972a; Stanton 1973; Roberts and Paul 1978; Vincent et al. 1991; Stanton et al. 1997; Chiswell and Booth 1999; Chiswell and Schiel 2001; Chiswell 2005; Reynolds-Fleming and Fleming 2005). Eddies can retain larvae enabling local recruitment (Chiswell and Roemmich 1998). Upwelling can sweep larvae offshore preventing dispersal along coasts (Veale 2007). Although river plumes have not yet been implicated in the structuring of New Zealand coastal marine invertebrates, they could potentially limit dispersal by increasing the amount of suspended sediment in the water, which has been linked with increased mortality of marine invertebrate larvae (Phillips and Shima 2006).

Genetic splits around Cook Strait

Early genetic studies of New Zealand marine fauna were economically motivated and aimed at providing information for fisheries stock assessment (e.g., Smith 1988; Smith et al. 1989; Smith and Benson 1997; Smith et al. 1997). These early analyses indicated genetic heterogeneity among sampling sites for the orange roughy (Smith and Benson 1997; Smith et al. 1997), the greenshell mussel (Smith 1988), and the surf clam (Smith et al. 1989). Although

the studies of greenshell mussels and surf clams consisted of sampling several locations spread across New Zealand, they identified a population genetic trend that would be more thoroughly explored in later studies: specifically, a north-south split around the Cook Strait region (Figure 2.1).

A series of papers stemming from Jonathan Gardner's laboratory examined population genetics in greenshell mussels, *Perna canaliculus*, using multiple genetic markers (allozymes, Apte and Gardner 2001; mitochondrial DNA, Apte and Gardner 2002; RAPDs, Star et al. 2003). Although the results from allozymes suggested panmixia (Apte and Gardner 2001), Apte and Gardner (2002) defined two different groups of *P. canaliculus* based on haplotype frequency differences in mitochondrial NADH dehydrogenase subunit IV (NADHIV). Sampling localities in the North Island and north of the South Island were distinct from the remaining South Island populations leading Apte and Gardner (2002) to suggest a phylogeographic break at the southern limit of the Cook Strait region (about 42° S; Figure 2.1). Similarly, a follow-up study using randomly amplified polymorphic DNA (RAPD) supported the structure identified in the mitochondrial DNA (Star et al. 2003). As Apte and Gardner (2002) suggested a north-south split, Sponer and Roy (2002) independently identified a north-south split both among major lineages and within a widespread lineage of the brittle star *Amphipolis squamata*; however, this split was suggested to occur between the North and South Islands.

The possibility of a breakpoint centered in the Cook Strait region has provided an a priori hypothesis for further studies of New Zealand marine invertebrates. Using mitochondrial DNA, Waters and Roy (2004) tested 1) a north-south split at Cook Strait narrows and 2) a north-south split at 42° S in another brittle star, *Patriella regularis*. The genetic structure observed supported both hypotheses. Adding three more sampling locations and more individuals from previous locations, Ayers and Waters (2005) sought to "pinpoint" the disjunction. Again, disjunctions at both the Cook Strait narrows and at 42° S were significant. Ayers and Waters (2005) further tested for differences by contrasting mitochondrial DNA between sampling locations in central New Zealand and sampling locations in the south of the South Island. This test indicated a significant difference between the regions. In contrast, no significant difference was found when sampling locations from central New Zealand were compared to sampling locations from the North Island. As a result, they concluded the disjunction lay south of Cook Strait.

Goldstien et al. (2006b) also used the north-south split to examine patterns of differentiation in coastal limpets, *Cellana* spp., but unlike the above studies, they addressed

the hypotheses from a comparative phylogeographic approach. Whereas previous results could have resulted from species-specific attributes or marker choice, Goldstein et al. (2006b) examined mitochondrial cytochrome b of three closely related intertidal limpets. The limpets are broadcast spawners and have a short larval duration (3–11 days), yet differ in spawning time. All three species showed a north-south split, with two species having a genetic disjunction across the southern Cook Strait region. To help resolve the location of the barrier, Goldstien et al. (2006b) sampled limpets from Cape Campbell (Figure 2.1); however, this only complicated the conclusions and suggested that genetic connectivity across Cook Strait varies by species. Cape Campbell samples were more similar to the North Island samples for *C. radians* but were more similar to South Island samples for *C. ornata*. Furthermore, the *C. ornata* data did not support either a strict Cook Strait barrier or a strict 42° S barrier because a Marlborough Sounds sample grouped with the North Island while the Cape Campbell samples grouped with the South Island.

Like *C. ornata*, genetic structure of snakeskin chiton, *Sypharochiton pelliserpentis*, supported a north-south split but neither a Cook Strait proper split nor a 42° S split best explained the structure (Veale 2007). Two populations, Kaiteriteri and Ocean Beach, grouped with the North Island samples, while Whareriki and Ward Beach (above 42° S) grouped with the South Island samples (Figure 2.1, Veale 2007). Based on his study and reexamining the above literature, Veale (2007) places a phylogeographic boundary between Cape Campbell on the east coast and Farewell Spit on the west coast (Figure 2.1).

Cook Strait as a barrier

The 23 km wide Cook Strait narrows only became a contemporary feature of New Zealand's coastal oceanography between 10,000–5,000 ya with the rise of sea levels at the end of the last ice age (Stevens et al. 1995). The hydrography around the Cook Strait region is complex. The region is relatively shallow with 96% of the area less than 200 m deep (Harris 1990). The bathymetry of the region includes smooth sea floor in the north and west that is interrupted by canyons in the east (Harris 1990). The sea floor contains broad expanses of sand and mud, and shorter stretches of pebbles and coarse sand in high stress areas. Strong tidal flows in some areas cause erosion and inhibit deposition (Harris 1990).

Three currents converge on Cook Strait: the D'Urville Current from the west, the Southland Current from the southeast, and the East Cape Current from the northeast (Figure 1.4 and Figure 2.1, Heath 1970). The product of this convergence is a net northwest flow (Heath 1985). The waters entering the Cook Strait region have different properties such that

warmer more saline water is in the northwest and northeast, cooler less saline water is in the southwest, and even cooler water is in the southeast (Heath 1970; Harris 1990; Vincent et al. 1991). Temperature and salinity within the Cook Strait region are further altered by river discharge, which changes seasonally (Harris 1990).

The different waters within the Cook Strait region potentially mix via irregular eddies (Vincent et al. 1991) and tides (Harris 1990). The tides around New Zealand are amphidromic. As a result, when tides are high in the west, tides are low in the east. This coupled with short tidal changes within Cook Strait proper produces strong tidal flows and a large amount of tidal mixing (Heath 1978, Harris 1990; Hume et al. 1992). Wind activity is correlated with tidal flow in Cook Strait proper (Heath 1986), current flow entering the Cook Strait region (Harris 1990), and the presence and strength of features such as eddies and upwelling in the Cook Strait region (Heath 1972b; Barnes 1985; Bowman et al. 1983; Bradford et al. 1986; Murdoch et al. 1990; Shirtcliffe et al. 1990; Vincent et al. 1991; Bradford-Grieve et al. 1993).

How Cook Strait acts as a barrier is unknown. The hydrography of the Cook Strait region may present a contemporary physical barrier to gene flow. Through comparing the north-south split in *S. pelliserpentis* (Veale 2007), *C. ornata* (Goldstien et al. 2006b), *C. radians* (Goldstien et al. 2006b), *P. regularis* (Ayers and Waters 2005; Waters and Roy 2004), and *P. canaliculus* (Apte and Gardner 2002; Star et al. 2003), Veale (2007) located the north-south split to lie on the northern coast of the south island: specifically, samples from locations between Farewell Spit on the west coast and Cape Campbell on the east coast were more similar to samples from the North Island. The areas separating these samples from the remaining South Island samples corresponded to areas of reported upwelling, Cape Farewell/Farewell Spit in the west and Cloudy and Clifford Bays in the east (Figure 2.1, Veale 2007 and references therein).

Upwelling involves the displacement of warmer surface water with cool, dense subsurface waters and could potentially limit gene flow by limiting dispersal. The displaced surface water can travel at speeds that are 10–100 times faster than larvae swimming at ≤ 0.1 cm s^{-1} (Shanks and Brink 2005). As a result, larvae may be swept offshore and away from suitable habitat for settlement (Roughgarden et al. 1988). On the other hand if upwelling is intermittent, periods of relaxation and downwelling can move larvae onshore (Farrell et al. 1991).

Unfortunately whether upwelling really limits larval dispersal is arguable. Few studies have directly examined the effects of upwelling on larval dispersal, and their results were

variable. For instance, Roughgarden et al. (1988) reported that barnacle (*Balanus glandula*) larvae were found offshore following upwelling in central California and suggested that loss of larvae offshore might be responsible for the low levels of recruitment during the three weeks following upwelling events.

On the contrary, Poulin et al. (2002) and Shanks and Brink (2005) found larvae distributed near the coast despite upwelling events. During upwelling events, *Concholepas concholepas* was present in recently upwelled water located between the upwelling front and the Chilean shore (Poulin et al. 2002). Poulin et al. (2002) suggested that downward vertical movement of *C. concholepas* larvae at night relocated them into the cold upwelling water moving toward the shore. Shanks and Brink (2005) also found the vertical positioning of the larvae to be an important determinant of the effects of upwelling. When slow swimming larvae of intertidal bivalves, *Spisula solidissima* and *Ensis directus*, were distributed in the deeper colder water (below the thermocline), they were transported offshore during downwelling episodes and onshore during upwelling episodes (Shanks and Brink 2005). In contrast to *S. solidissima* and *E. directus*, *Tellina* spp. and *Mulinia lateralis* remained near shore during both upwelling and downwelling events (Shanks and Brink 2005). Shanks and Brink (2005) proposed a model in which larvae maintain a preferred depth by moving vertically against upwelling or downwelling water, and if larvae got caught in a cross shelf current then eventually vertical movement would position them in water that would carry them back to shore.

Although upwelling in the greater Cook Strait region has been linked to changes in macrozooplankton assemblages (Bradford-Grieve et al. 1993) and corresponds with the putative north-south split (Veale 2007), its role in dispersal and gene flow remains poorly understood, particularly in terms of larval biology and migration intensity of New Zealand species. The magnitude of genetic structure around New Zealand varies according to species. Based on different regions of the mitochondrial genome, species like *C. ornata* and *S. pelliserpentis* have large Φ_{ST} values (0.829, Goldstien et al. 2006b, and 0.45, Veale 2007, respectively) compared to *C. radians* (Φ_{ST} = 0.142, Goldstien et al. 2006b), *P. regularis* (Φ_{ST} = 0.072, Waters and Roy 2004), and *P. canaliculus* (Φ_{ST} = 0.162, Apte and Gardner 2002). Although Φ_{ST} values are not directly comparable, the differences between these values are large.

Variable amounts of Φ_{ST} may be due to larval behavior and life history characteristics (Hedgecock 1986; Bohonak 1999). Goldstien (2005) and Veale (2007) found no correlation between larval longevity and overall magnitude of population structure. Potentially, other

larval characteristics like behavior could be factors but have not been assessed. Instead, Φ_{ST} variation was correlated with interspecific differences such as spawning time (Goldstien 2005; Veale 2007). At least some of the upwelling in the Cook Strait region is seasonal (Bradford-Grieve et al. 1993), therefore the upwelling should be a stronger barrier to gene flow only if it coincides with the time of spawning. Spawning times for *C. ornata* and *S. pelliserpentis* occurred during the upwelling season for the western Cook Strait region leading Veale (2007) to suggest this as the cause for the larger Φ_{ST} values. Yet the upwelling in the eastern region between Cloudy Bay and Clifford Bay is persistent regardless of wind (Bradford et al. 1986). For species that do not spawn during the western Cook Strait upwelling season, no evidence supporting asymmetric gene flow exists (e.g., *C. radians*, Goldstien et al. 2006b; *P. canaliculus*, Apte and Gardner 2002; *P. regularis*, Ayers and Waters 2004). Furthermore, upwelling may limit dispersal north and south but it does not explain other trends occurring in the region. For instance, *S. pelliserpentis* haplogroup C was common in two South Island locations, which were north of the upwelling zones; however, haplogroup C was absent from the North Island (Veale 2007).

The upwelling hypothesis invokes contemporary hydrology, whereas the genetic structuring around Cook Strait may, instead, reflect past conditions. For example, Stevens and Hogg (2004) suggested that the large genetic distance between samples of amphipods from Cook Strait and the North Island was due to isolation of the Cook Strait sample in a "lake" prior to the development of Cook Strait rather than a contemporary process. The upwelling hypothesis does not reconcile well with attempts to date the north-south split using the genetic data for *P. canaliculus* and *C. ornata* (Apte and Gardner 2002; Goldstien et al. 2006b, respectively). For both species, the split probably existed for much longer than the contemporary upwelling. Sequence divergence between an ancestral haplotype and a unique South Island haplotype for *P. canaliculus* indicates a divergence time of 1.3 mya prior to postglacial upwelling (data from Apte and Gardner 2002 as recalculated in Goldstien et al. 2006b). Nested clade analysis on *P. canaliculus* could not distinguish between historical range expansion and contemporary restricted gene flow (Apte and Gardner 2002). The most recent common ancestor for *C. ornata* dates to 0.24–0.3 mya, a time period with inconclusive paleoceanography (Goldstien et al. 2006b).

H. iris genetic structure

Dollimore (1977) compared esterases between three sites on the North Island and two sites on the south island and found no distinction among sites. Using phosphoglucomutase,

Frusin (1982) found *H. iris* from the Chatham Islands to be different from Wellington and Kaikoura samples. Driven by possibilities of stock enhancement and the paucity of previous data, the New Zealand Ministry of Fisheries is now funding genetic studies of New Zealand abalone stocks. A pilot study using 630 bp of the ATPase8–ATPase6 region of mitochondrial DNA showed weak differentiation between the Chatham island and three mainland sites (Stewart Island, Great Barrier Island, and South Taranaki, Smith and McVeagh 2006). Smith and McVeagh (2006) admitted that mainland structure may have been missed due to sampling few locations and small sample sizes, only 10–11 individuals per site.

To supplement Smith and McVeagh's (2006) preliminary findings, this chapter analyzed variation identified in two regions of the mitochondrial genome (COI and ATPase8–ATPase6) for *H. iris* sampled from 25 locations around New Zealand to 1) identify if genetic structure exists within *H. iris*, 2) if structure does exist, determine the whether the pattern is consistent with a genetic split across Cook Strait.

MATERIALS AND METHODS

Marker choice

Marker choice in genetic studies is important because marker characteristics (e.g., mode of inheritance and mutation rates) will influence the interpretation of the results. Mitochondrial DNA was used in this study due to its relative ease of amplification (large copy number compared to nuclear DNA), its haploid state, lack of recombination (Birky 2001; although see Innan and Nordborg 2003), and maternal inheritance (Birky 1995; although see Zouros 2000; and Kondo et al. 1990). The haploid genome and maternal inheritance mean that mtDNA has a lower effective population size (N_e) than nuclear DNA (Hartl and Clark 1997; although see Ballard and Whitlock 2004). The mitochondrial genome also has an elevated rate of evolution compared to the nuclear genome (Brown et al. 1979; Denver et al. 2000). Lower effective population sizes and an increased evolutionary rate mean that mitochondrial DNA will achieve equilibrium faster than nuclear DNA and may be a better signal of population level processes (e.g., gene flow, bottlenecks, and founder effects, Moritz et al. 1987).

A 459 bp region of mitochondrial cytochrome oxidase I (mtCOI) was amplified because Metz et al. (1998b) had already successfully amplified this region in *H. iris* DNA, and intraspecific variation in mtCOI has been reported in other abalone species, such as *H. rufescens* (Burton and Tegner 2000) and *H. cracherodii* (Hamm and Burton 2000). Preliminary screening of 20 adult *H. iris* with Metz et al.'s (1998b) mtCOI primers confirmed

intraspecific variation. With publication of the full *H. rubra* mitochondrial genome (Maynard et al. 2005), an additional 596 bp of a region spanning from ATPase8–ATPase6 was amplified. Smith and McVeigh (2006) had already found this region to be variable in *H. iris*.

The mtCOI and the ATPase8–ATPase6 regions are separated by 1526 bp, which includes mtCOI and mtCOII in *H. rubra* (Figure 2.2, Maynard et al. 2005). Maynard et al. (2005) determined ATPase8–ATPase6 region to be highly variable and the mtCOI region to be more conserved. They also identified a putative control region (a highly variable region that is frequently employed in phylogeographic studies); however, this control region contained large repetitive AT tracts making it difficult to amplify and sequence.

MtCOI, ATPase6, and ATPase8 encode for peptides that make up mitochondrial transmembrane proteins involved in the electron transport chain and ATP synthesis. Specifically, the mtCOI subunit forms part of the cytochrome oxidase complex in the electron transport chain, while the ATPase subunits help form the F_0 complex in F_0 F_1 ATPase complex, which uses the proton motive force created by the electron transport chain to catalyze ATP hydrolysis (Lodish et al. 1995). Given the functional importance of mitochondria and ATP synthesis, mounting evidence suggests that mitochondrial DNA are subject to direct and indirect selection (Ballard and Whitlock 2004 and references therein).

Samples

Haliotis iris is distributed ubiquitously throughout the coastal waters of New Zealand. Consequently, in order to undertake a robust study of the genetic diversity of this species, *H. iris* was sampled as widely and extensively from this distribution as practical with the resources and the time available. The sampling strategy was a classical cost/power compromise for examining genetic diversity under financial constraints that focused on sampling more populations less intensively rather than sampling a few populations intensively (Pons and Chaouche 1995; Pons and Petit 1995).

Foot or mantle tissue was collected from between 13–24 abalone at 25 sites around New Zealand (Figure 2.3, Appendix 1). Individuals were collected by a variety of means, i.e. confiscated illegal catch, commercial catch, recreational catch, and scientific catch, making different tissue available for different samples. All attempts were made to collect samples from mature abalone; however, not all abalone were larger than legal minimum size (12.5 cm).

DNA extraction, PCR amplification, and sequencing

After multiple trials of different extraction techniques, the most consistent and malleable DNA was produced using Qiagen's DNEasy® Blood & Tissue Kit. The kit was used to extract abalone DNA from 24 of the 25 sites. Abalone DNA from the remaining site (OPT) was extracted using a modified LiCl protocol (Gemmell and Akiyama 1996). Metz et al.'s (1998b) mtCOI primers (F1 and R1) amplified 581 bp. However due to inconsistent amplifications, internal mtCOI primers (mtCOI_F2, mtCOI_R2) were designed with Primer3 (Rozen and Skaletsky 2000). MtCOI_F2 and mtCOI_R2 amplified 540 bp (Figure 2.3). Primers COIIcons-F and H22-R1 (Maynard et al. 2005) consistently amplified 723 bp of the ATPase8–ATPase6 region.

Both mtCOI and ATPase8–ATPase6 were amplified in a 25 μL reaction volume containing 1–40 ng of genomic DNA, 200 μM dNTPs, 0.4 μM of each primer, 1.5 mM $MgCl_2$, 1X NH_4 Reaction Buffer (160mM $(NH_4)_2SO_4$, 670 mM Tris-HCl (pH 8.8 at 25 °C), and 0.1% Tween-20), and 0.5 units BIOTAQ™ (Bioline USA, Inc.). A negative control was included in all PCRs. Thermal cycling parameters were an initial denaturation at 96 °C for two minutes, 35 cycles of 96 °C/20 s, 55 (mtCOI) and 60 (ATPase8–ATPase6) °C/30 s, 72 °C /30 s (mtCOI) and 45 s (ATPase8–ATPase6), and a final cycle elongation step at 72 °C for seven minutes. Amplification products were checked on 1% agarose gels against a Lambda DNA/*EcoRI+HindIII* ladder and were visualized under UV following staining with ethidium bromide (0.5 μg/mL). Successful amplifications were purified according to manufacturer's instructions using either a vacuum method with Eppendorf Perfectprep® PCR Cleanup 96 plates or a centrifugation method with PALL® AcroPrep™ 96 Filter Plate (Omega 30k).

Purified amplicons were subject to direct sequencing with ABI Prism® Big Dye® Terminator v. 3.1 Cycle Sequencing Kit (Applied Biosystems, Inc.) as per the manufacturer's instructions but at 1/8th the suggested volume. Sequence products were purified using Sephadex™ GS-50 gel filtration (Amersham Bioscience) and run on an ABI3100 Genetic Analyzer (Applied Biosystems, Inc.) at the University of Canterbury. Sequences were edited with Sequencher™ 4.2.2 (Gene Codes Corporation). Sequence alignment was conducted by hand using Se-Al v2.0a11 (Rambaut 2002), and all variable sites were confirmed by visual inspection of chromatograms. A total of 459 bp of mtCOI and 596 bp of ATPase8–ATPase6 were obtained from 477 out of 478 individuals.

Population genetic analyses

Due to the linked nature and correlated evolution of mitochondrial DNA, the two regions were concatenated giving a total of 1055 bp for analyses. In order to assess sequence

variation within samples, standard molecular indices were calculated. The number of polymorphic sites, haplotype diversity (*h*, Nei 1987), and nucleotide diversity (π, Nei 1987) were computed for each location as well as for a priori groups, listed in Table 2.2, using Arlequin 3.1 (Excoffier et al. 2005). These are all measures of polymorphism with added benefits over simply the number of haplotypes. The number of polymorphic sites accounts for the number of differences between sequences and is applicable when all the observed sequences are different, but is dependent on sample size (Nei 1987; Nei and Kumar 2000). In contrast, both haplotype diversity and nucleotide diversity are measures of polymorphism independent of sample size (Nei 1987; Nei and Kumar 2000). Specifically, haplotype diversity or average gene diversity is the probability that two haplotypes from a sample are different, while nucleotide diversity is the average number of nucleotide differences per site and is independent of sequence length as well as sample size (Nei 1987; Nei and Kumar 2000).

Haplotype relationships

To quantitatively assess similarity and differences among haplotypes, percent divergence between haplotype pairs was calculated using maximum likelihood settings for the concatenated sequences in PAUP*4.10b10 (Swofford 1998). Maximum likelihood parameters were established for mitochondrial regions (both separately and concatenated) in MODELTEST 3.7 (Posada and Crandall 1998). According to Akaike information criterion (AIC, Posada and Buckley 2004), the most appropriate models of sequence evolution were TrN+I (proportion of invariable sites I = 0.7825, Tamura and Nei 1993) for mtCOI, GTR+I (proportion of invariable sites I = 0.5366, Tavaré 1986) for ATPase8 region, and GTR+G (gamma distribution shape parameter G = 0.1473, Tavaré 1986) for the concatenated sequences. The Tamura and Nei (TrN) model uses variable base frequencies, equal transversion frequencies and variable transition frequencies to describe sequence evolution, while the General Time Reversible (GTR) model uses variable base frequencies and a symmetrical substitution matrix to describe sequence evolution (Posada and Crandall 1998; Felsenstein 2004).

To visually assess similarity and differences among haplotypes, haplotype networks were constructed. When working with intraspecific gene genealogies, Posada and Crandall (2001) argue for the use of a network approach. Relationships between haplotypes are inferred with three frequently used network-building algorithms: median-spanning (Excoffier and Smouse 1994, implemented in Arlequin 3.1, Excoffier et al. 2005), median-joining (Bandelt et al. 1999, implemented in Network 4.2.0.1, Fluxus Technology Ltd.), and

statistical parsimony (Templeton et al. 1992, implemented in TCS Clement et al. 2000). Cassens et al.'s (2003) initial assessment of these three different algorithms demonstrated that they generate different networks from the same data indicating systematic differences among the algorithms. Cassens et al.'s (2005) comparisons of the algorithms with simulated data showed that minimum spanning networks on average required more mutational steps to resolve a network and that median-joining networks produced more accurate networks when internal node haplotypes were missing.

Although median-joining may be the most accurate, a minimum spanning and a statistical parsimony network were also constructed because the pairwise differences used in AMOVAs (discussed below) are similar to patristic distances in minimum spanning networks (Excoffier and Smouse 1994; Excoffier 2003), and statistical parsimony networks are traditionally used in Nested Clade Phylogeographic Analysis (Templeton 1998).

Cluster analysis

Different patterns of genetic structuring have been identified for New Zealand coastal invertebrates (Table 2.1), so potential patterns of genetic structure among *H. iris* samples were first explored using cluster analyses. Net genetic distances (Nei and Li 1979) between each pair of sampling locations were measured in Arlequin 3.1 (Excoffier et al. 2005). The net genetic distance (d_A) is estimated by taking the average number of nucleotide substitutions between haplotypes from two populations (d_{xy}) and correcting for the average number of nucleotide substitutions that existed when the populations split (Nei 1987). No reason existed a priori to assume a past bottleneck (Hedrick 1999), therefore the d_{xy} values were used in clustering analyses. Relationships in genetic differentiation (d_{xy}) among sampling locations were visualized using metric multidimensional scaling computed in R version 2.6.1 (Team 2007), and the stress was calculated according to Venables and Ripley (1999 p. 333). As a result of a large stress value (0.48), further clustering of d_{xy} was performed with the minimum evolution principle and no assumption of constant rates using the neighbor joining algorithm (Nei and Kumar 2000; Felsenstein 2004) implemented in MEGA4 (Tamura et al. 2007).

MDS and the neighbor joining algorithms used above visualize genetic structure based solely on genetic distances and do not incorporate the spatial relationship between samples. In contrast, a spatial analysis of molecular variance (SAMOVA) is essentially a clustering approach that assigns samples from geographically adjacent sites to a predefined number of groups in order to maximize the between group genetic variance determined using AMOVAs (Dupanloup et al. 2002). SAMOVA is a fairly recent approach applicable to both genotypic and haplotypic data. SAMOVAs were performed in SAMOVA 1.0 (Dupanloup et al. 2002).

SAMOVA 1.0 can only read geographical input files that contain x, y coordinate data, which does not accurately reflect distances between locations separated in one dimension or linear space (i.e., coastal distances).

Hypothesis testing

The above cluster analyses tried to determine structure a posteriori. Because a considerable amount of literature exists on New Zealand coastal invertebrates, samples were divided into a priori groups (Table 2.2) based on 1) a mainland-Chatham Islands genetic split (Smith and McVeagh 2005) and 2) a north-south genetic split (Apte and Gardner 2002; Star et al. 2003; Waters and Roy 2004; Ayers and Waters 2005; Goldstien et al. 2006b; Veale 2007). A priori statistics are desirable because they are more powerful than a posteriori statistics.

To test these a priori groups, analyses of molecular variance (AMOVAs), based on the number of pairwise differences, were employed (Excoffier et al. 1992; Excoffier 2003). AMOVAs were implemented in Arlequin 3.1 (Excoffier et al. 2005), and significance tests consisted of using 16002 permutations. AMOVAs index genetic variation with Φ statistics and partition the total variation into three components: Φ_{ST} is the correlation of random haplotypes within demes relative to random pairs of haplotypes drawn from the whole species; Φ_{SC} is the correlation random haplotypes within demes relative to the correlation of random pairs of haplotypes drawn from within defined groups; Φ_{CT} is the correlation of random haplotypes within groups relative to random pairs of haplotypes drawn from the whole species (Excoffier et al. 1992; Excoffier 2003). Here, sampling locations are equivalent to demes. If Φ_{CT} is significant, then the proposed grouping accounts for more of the total variation than expected by chance alone. When multiple structures were found to be significant, these were evaluated according to Φ_{CT} and Φ_{SC} values: the scenario that resulted in the highest Φ_{CT} and therefore lowest Φ_{SC} value ($(1-\Phi_{ST}) = (1-\Phi_{SC})(1-\Phi_{CT})$ assuming Φ_{ST} remain constant) was considered to fit the data better.

Linking geographic and genetic distances

The relationship between geography and haplotype was examined using nested clade (Templeton et al. 1995; Templeton 1998) and isolation by distance approaches (Mantel tests, Mantel 1967). Each approach requires geographic distance between sampling locations, and coastal distances between locations were determined using the geographical information system, ArcMap™ 9.1 (Environmental Systems Research Institute, Inc.) and Google™ Earth 4.3. Unlike SAMOVAs, both nested clade and IBD programs can incorporate distance matrices.

Nested clade phylogeographical analysis (NCPA) was originally performed to search for associations between phenotypes and genotypes (Templeton et al. 1987; Templeton et al. 1992). NCPA was adapted for statistically testing the association between geographic distribution and haplotypes (Templeton et al. 1995; Templeton 1998) and has, since, been a widely used tool in phylogeography (Templeton 2004). In addition to simply testing isolation by distance, NCPA attempts to distinguish between historical and contemporary processes (e.g., continuous range expansion and population fragmentation). The analysis involves calculating the geographical spread of clade members relative to the mean location (clade distance, D_c) and the geographical spread of clade members relative to mean location of members of the nesting clade (nested clade distance, D_n). The significance of D_c and D_n are tested through permutations and used to infer no geographical association versus potential population processes (Posada et al. 2006).

Despite the inclusion of NCPA in many studies (e.g., Uthicke and Benzie 2003; Imron et al. 2007; Richards et al. 2007), researchers are compelled to corroborate their findings with IBD tests and mismatch distributions; they distrust NCPA and with good cause (most recently debated in Petit 2008a, b; Garrick et al. 2008). First, NCPA has a degree of circularity. Networks should not contain cycles, and cycles are typically broken, based on the coalescent, according to Crandall and Templeton (1993): 1) rare haplotypes are more likely found at tip positions, while common haplotypes are more likely found at interior positions, and 2) singletons are more likely connected to haplotypes from the same location than haplotypes from different locations. These assumptions will minimize geographical spread within clades and maximize the geographical spread between clades. Second, the final step requires following an inference key derived from "known" instances of a process. Knowles and Maddison (2002) criticized the a posteriori findings of NCPA with a small-scale simulation study. In defense, Templeton (2004) claimed that NCPA performs well in instances of a priori knowledge. Counter to Templeton's (2004) examination of real data, Panchal and Beaumont's (2007) simulation study found NCPA generated a high frequency of false positives, particularly for inferences of restricted gene flow with isolation by difference and contiguous range expansion.

Nonetheless, NCPA was conducted here to help interpret haplotype networks. Although like other studies, additional tests of isolation by distance and range expansion were also conducted. Cycles on the statistical parsimony network, generated above, were broken based on Crandall and Templeton (1993). The network was nested according to Templeton and Sing (1993) in AneCA v1.2 (Panchal 2007). D_c, D_n, and significances were calculated

with GeoDis 2.5 (Posada et al. 2006), and interpretation of significant findings used the 11 November 2005 inference key provided with GeoDis 2.5 (Posada et al. 2006).

Isolation by distance was examined using a Mantel test (Mantel 1967) implemented in Arlequin 3.1 (Excoffier et al. 2005). Correlations of coastal distances and pairwise Φ_{ST} values (Reynolds et al. 1983) for each pair of sampling locations were calculated and tested for significance with 10,000 permutations. Note that linearized Φ_{ST} values were not used because they assume populations have been isolated without migrants (Slatkin 1991). The sharing of haplotypes among sampling locations suggested migration has occurred, violating this assumption. Mantel tests were performed for all sampling locations and for groups defined in (Table 2.2).

Mismatch distributions

In addition to NCPA and IBD, information about processes affecting genetic structure can also be gleaned from using mismatch distributions. The distribution of pairwise differences in a sample of DNA sequences depends on the demographic history of the population (Slatkin and Hudson 1991). After a recent demographic expansion, the distribution of pairwise differences will approximate the Poisson distribution; instead for a population of constant size, the distribution of pairwise differences tends to be multimodal and erratic (Slatkin and Hudson 1991; Rogers and Harpending 1992). The observed numbers of pairwise differences between haplotypes were compared with simulated data under two models of expansion: pure demographic expansion (Slatkin and Hudson 1991; Rogers and Harpending 1992) and spatial expansion with migration (Ray et al. 2003; Excoffier 2004). Comparisons were made for collection sites, a priori groups listed in Table 2.2, and NCPA clades 4-1 and 4-2 (implemented in Arlequin 3.1, Excoffier et al. 2005). Significant deviations from the models were assessed with an ad hoc statistic, raggedness (Harpending et al. 1993; Harpending 1994), and using sum of squared deviations.

Neutrality tests

Tajima's *D* and Fu's F_s were used to assess sequence neutrality. These tests require only DNA polymorphism data. Tajima's *D* examines the relationship between the number of segregating sites and nucleotide diversity (Tajima 1989). In mutation-drift equilibrium, the number of segregating sites and nucleotide diversity estimate the same value of θ ($\theta = 4N_e\mu$). When deleterious mutations are present, the number of segregating sites increases, while nucleotide diversity remains the same (Tajima 1989). When mutant site frequencies increase (e.g., due to balancing selection), nucleotide diversity increases, while the number of

segregating sites remains the same (Nei and Kumar 2000). Tajima's *D*, therefore, measures the differences between the number of segregating sites and nucleotide diversity to infer selection. Importantly, selection can be inferred only if the population is at equilibrium; if not, then other evolutionary factors (e.g., a bottleneck) may be present.

Fu's F_s examines the probability of the number of recent mutations given a value θ (Fu 1997). Whereas excesses of old or more common alleles indicate population subdivision, population shrinkage, and balancing selection, excesses of recent mutations or rare alleles indicate population growth, hitchhiking, and background selection (Fu 1997). With excesses of recent mutations, θ estimated from the mean number of nucleotide differences between two sequences will be smaller than θ estimated from the number of alleles (Fu 1997). F_s values tend to be negative when recent mutations are in excess. (Note F_s values are significant at the 5% level if the p-value is below 0.02 and not 0.05, Fu 1997). Tajima's *D* and Fu's F_s were calculated in Arlequin 3.1 (Excoffier et al. 2005) for each sampling location and a priori groups of localities defined Table 2.2.

Bonferroni corrections

Bonferroni corrections control for experiment-wise error, which results from multiple tests on the same data set. Although each pairwise Φ_{ST} comparison examined a different pair of samples, each sample was used in multiple comparisons. Therefore, standard Bonferroni corrections were presented for pairwise Φ_{ST} comparisons. Note, Bonferroni corrections were conservative and alternative corrections (see Rice 1989; Narum 2006) might be more applicable but were not presented here.

RESULTS

A 459 bp fragment of mtCOI and a 596 bp fragment of the ATPase8–ATPase6 were amplified in 477 *H. iris* from 25 locations around New Zealand. The mtCOI fragment and the ATPase8–ATPase6 fragment corresponded to base pairs 3504–3612 and base pairs 5582–6178, respectively, in the *H. rubra* mitochondrial genome (ACCN: NC_0059400). As separate fragments, ATPase8–ATPase6 was more variable than the mtCOI: it had a larger number of polymorphic sites, higher haplotype diversity, and greater nucleotide diversity (Table 2.3). The mtCOI contained no indels and 13 polymorphic sites that were parsimony informative, while ATPase8–ATPase6 contained two indels and 39 parsimony informative sites. The majority of mutating sites had two variants, while one mtCOI site and four ATPase8–ATPase6 sites had three variants.

Haplotypes: Relationships and distributions

As concatenated sequences (bp = 1055), a total of 132 haplotypes were identified from 119 polymorphic sites (113 transitions and 9 tranversions), 52 of which were parsimony informative (Table 2.3; Appendices 2 and 3). Percent pairwise divergence between haplotypes calculated using a maximum likelihood approach ranged from 0.0948–1.3282%. The haplotype diversity of all individuals grouped as single population was 0.8990±0.0081; otherwise haplotype diversity ranged from 0.5824 (DSD)–0.9810 (SPB) when individuals were grouped according to sampling locations (Table 2.4). The nucleotide diversity of all individuals treated as a single population was 0.003827±0.002117, while nucleotide diversity ranged from 0.001323 (MTB)–0.005478 (IHM) within sampling locations (Table 2.4).

Only 23 haplotypes were shared among locations, the remaining 109 haplotypes were private (Figure 2.4). Four haplotypes (numbered 8, 10, 17, and 18 in Figure 2.5–2.7) were identified in more than 40 individuals. Haplotype 8 was found in all locations except GLN. Haplotype 10 was the most prevalent and found in all locations except the Chatham Islands (OCH). Haplotype 17 was absent in all North Island locations except OLB and MAT and present in OCH and all South Island locations except MTB. Haplotype 18 was also missing from all but four (EAI, OLB, MAT, and WLG) North Island locations and present in OCH and all the South Island locations. More private haplotypes were found in the North Island (74 haplotypes) than in the South Island (32 haplotypes). The concentration of more private haplotypes in northern locations was also indicated with higher haplotype diversities in these locations (Table 2.4). The only locations without any private haplotypes were CCB, DSD, and WST from the South Island.

In general, haplotypes were closely related (Figures 2.5–2.7) as expected according to the low nucleotide diversity (Table 2.3). The three network algorithms consistently reported three common haplotypes, 8, 10, and 18, that were only one or two mutations apart (Figures 2.5–2.7). Stemming from these main haplotypes were many singleton haplotypes creating star-like formations. The fourth main haplotype, 17, was seven to nine mutations away from the other three common haplotypes. It was located in a web-like region that contained rare and missing haplotypes. This web-like region also contained the most discrepancies among the networks with many more cycles present in the minimum spanning network.

Cluster analyses

Significant population genetic structure existed among all samples (Φ_{ST} = 0.04453, p = 0.000). Cluster techniques were applied to visually inspect the data and identify patterns of

genetic structure or substructure. Multidimensional scaling (MDS) based on Nei's d_{xy} revealed no obvious groupings of the sampling locations (Figure 2.8). Instead, the sampling locations were ordered in a continuum with South Island toward a side and North Island locations lying toward the other side. However, a large stress value (0.48) indicated that the configuration in Figure 2.8 was a very poor representation of the data. Rather than increase the number of dimensions and, hence, the number of parameters, the neighbor joining method was used as an alternative clustering technique. The neighbor joining tree had a star-like topology with long branches leading to leaves and short internal branches between clusters (Figure 2.9). Three clusters were evident: 1) North Island samples and TCL from the South Island, 2) South Island samples and MAT and EAI from the North Island, and 3) South Island samples, Chatham Islands sample (OCH), and OLB from the North Island (Figure 2.9).

Unlike MDS and neighbor joining approaches, SAMOVAs included geographical data when grouping demes (Dupanloup et al. 2002). The Φ_{CT} estimate was maximized when the number of groups specified was two: 1) Chatham Islands and 2) North and South Islands (Table 2.5). This division explained 11.9% of the total variance, while the division of populations within groups explained 3.3% of the total variance. Further increases in the number of groups resulted in decreases in both Φ_{CT} and Φ_{SC} estimates. SAMOVAs excluding the OCH suggested that two groups were the best structure with IHM as one group and the remaining samples as the other group. IHM's separation was consistent with its distant position in the MDS and its long branch in the neighbor joining tree.

A priori hypothesis testing

AMOVAs were used to test the genetic structures proposed in Table 2.2. Smith and McVeagh's (2006) preliminary study of *H. iris* suggested that the Chatham Islands *H. iris* were distinct from the North and South Islands *H. iris*. An AMOVA comparing the Chatham Islands sample to the remaining North and South Island samples was significant and explained 11.9% of the variation (Table 2.6). To determine whether genetic structure existed between the North and South Island samples, an AMOVA excluding the Chatham Islands sample was performed. Grouping the 24 North and South Island samples as a single group produced a significant $\Phi_{ST} = 0.03815$ ($p = 0.000$). To test whether this genetic structure was related to Cook Strait region, further AMOVAs divided the samples into two different groups around Cook Strait. Both, splitting the samples across Cook Strait Narrows and splitting the samples according to the upwelling regions in Figure 2.1 produced significant Φ_{CT} indices (0.04521 and 0.04376, respectively). Between the two a priori structures, Φ_{CT} was maximized

when samples were split across Cook Strait Narrows (North Island vs. South Island). However, significant differentiation still occurred between samples within groups (Φ_{SC}). In all cases, the variance within sampling locations was very high and ranged from 84.8–95.6% of the total variance.

To better interpret AMOVA results, molecular indices for the groupings proposed in Table 2.2 and pairwise comparisons of Φ_{ST} were further inspected (Table 2.7 and 2.8). Noticeably, the number of haplotypes and the haplotype diversities were larger for groups that contained North Island samples (Table 2.7). In fact, haplotype diversities between northern and southern groups differed by more than two standard deviations. The pairwise Φ_{ST} showed that MTB and TIM (South Island), IHM (North Island), and Chatham Islands were the most divergent samples (Table 2.8). A greater number of significant comparisons (52 out of 143) occurred between North and South Island samples than between samples within either island (10 out of 55 for the North Island and 16 out of 78 for the South Island). After Bonferroni correction, only comparisons between South Island (MTB and TIM) and North Island (DBL, GLN, and IHM) samples were significant.

The pairwise Φ_{ST} indices for the two samples, PHD and TCL, which grouped either with the South Island or the North Island for the AMOVA groupings (Table 2.2), were not significantly different from any other samples after Bonferroni correction. Without Bonferroni correction, PHD was significantly different from two South Island samples (MTB and TIM), while TCL was significantly different from two South Island samples (MTB and TIM), one North Island sample (IHM), and the Chatham Islands sample (OCH).

Nested clade phylogeographic analysis

To better understand the structure of *H. iris*, nested clade phylogeographic analysis (NCPA) and isolation by distance tests were employed to examine process. NCPA revealed significant associations between haplotypes and geographic distributions. The total TCS network was nested in five steps (Figure 2.10). For the majority of the clades, panmixia could not be rejected. Eleven clades could be used with the inference key (Figure 2.10 and Table 2.9). Restricted gene flow and isolation by distance were inferred for five clades. Three clades seemed to suffer from a lack of genetic resolution or fine-scale sampling or both inhibiting differentiation between range expansion and restricted dispersal. Nothing could be inferred for the remaining three clades.

Mantel tests between genetic differentiation and geographic distance were performed for all sampling locations and for all groups proposed in Table 2.2 (Table 2.10). A Mantel test

across all locations was significant ($r^2 = 0.189$, $p = 0.047$). The Chatham Islands sample appeared to be responsible for this finding. Removing the Chatham Islands sample from the test reduced the correlation coefficient to 0.119 ($p = 0.126$), and a pattern of isolation by distance was no longer significant. Isolation by distance also occurred when PHD and TCL were grouped with the North Island samples ($r^2 = 0.237$, $p = 0.047$).

Mismatch distributions and neutrality tests

Further assessments of demographic processes were performed with mismatch distributions and neutrality tests. Mismatch distributions were computed for each sampling location, within each group specified in Table 2.2, and within clades 4-1 and 4-2 (Figure 2.10, Appendix 4). The models of demographic and spatial expansions were accepted for all mismatch distributions according to the raggedness values. Several sums of squared deviations were significant under the model of demographic expansion (CRW, MAT, OPT, WLG, WST, North Island, and North Island and north of South Island).

Tajima's *D* and Fu's F_s were significant for mtCOI, ATPase8–ATPase6, and concatenated fragments when all individuals were treated as a single group (Table 2.3). Samples EAI, TIM, and WLG samples had significantly negative *D* values (Table 2.4). TIM and WLG also had significantly negative F_s values. In addition, seven more North Island and three more South Island samples had significantly negative F_s values (Table 2.4). *D* was not significant for the a priori group consisting of all South Island samples except PHD and TCL. Otherwise, *D* and F_s were significant for all a priori groups concerning the North and South Islands (Table 2.7). Significant *D* and F_s implied that sequence polymorphisms deviated from neutrality as a result of population expansion or contraction and/or selection.

DISCUSSION

The *H. iris* samples collected from around New Zealand had highly significant albeit slight genetic structure. No striking patterns emerged from either the MDS or neighbor joining analyses, although with a few exceptions, northern and southern samples formed weak clusters. The largest Φ_{CT} values from the SAMOVAs suggested that two groups were present in New Zealand: Chatham Islands and mainland New Zealand. However, SAMOVAs did not identify structure between northern and southern samples consistent with those proposed in Table 2.2. All AMOVAs testing the structures proposed in Table 2.2 resulted in significant structure suggesting that the Chatham Islands sample was different from the mainland

samples and the northern samples were different from southern samples. Overall, northern samples typically had higher haplotype diversities than southern samples.

AMOVAs and SAMOVAs indicated that the amount of variation within sampling localities was large. *H. iris* mtDNA variation was characterized by high haplotype diversity resulting from a large number of rare haplotypes with few nucleotide differences (or low nucleotide diversity). Similarly high levels of haplotype diversity have been observed in other marine invertebrates (e.g., *Holothuria nobulis*, Uthicke and Benzie 2003; *Littorina keenae*, Lee and Boulding 2007); however, the haplotype networks in these instances resemble the shape of Clade 4-2 (Figure 2.10) and not the shape of Clade 4-1. AMOVAs for *H. nobulis* and *L. keenae* were not significant with the majority of variation attributed to variation within populations. Similarly, the majority of variation for *H. iris* was found within populations, yet unlike *H. nobulis* and *L. keenae* significant genetic structure was still evident in *H. iris*.

Large levels of genetic variation can accumulate in populations that are ancient, occupy a diversity of niches, or have an increased mutation rate, while large effective population sizes can reduce the loss of genetic diversity due to genetic drift. In the case of *H. iris***,** the presence of lots of rare haplotypes probably reflects a relatively recent population expansion. Although NCPA inferred mostly panmixia and restricted gene flow and only inferred population expansion for Clade 4-1, mismatch distributions consistently could not reject models of population expansion or spatial expansion. Significant Tajima's *D* and Fu's F_s also supported population expansion in *H. iris*.

The separation of the Chatham Islands sample from North and South Island samples was consistent with Smith and McVeagh's (2006) preliminary genetic study of *H. iris*. Smith and McVeagh (2006) found only seven haplotypes in 40 individuals from four locations with 631 bp of the ATPase8–ATPase6 region (0.175 haplotypes/individual). From 596 bp of the ATPase8–ATPase6 region sequenced here, 98 haplotypes were identified in 477 individuals (0.205 haplotypes/individual). Unfortunately, Smith and McVeagh's (2006) sequence data and haplotype diversities were not published, and they did not to explore variation in mtCOI. Randomization tests of the ATPase8–ATPase6 haplotype frequencies led Smith and McVeagh (2006) to conclude that differences existed between the Chatham Islands and the mainland *H. iris* and no differences existed among mainland sites. The inclusion of more sampling locations, more individuals, and/or an additional 459 bp of mtCOI has resulted in the detection of highly significant genetic structure among mainland samples.

Chatham Islands vs. North and South Island

In addition to the Smith and McVeagh (2006) study, the differentiation between Chatham Islands *H. iris* and mainland *H. iris* found here was also consistent with the few studies of New Zealand coastal marine invertebrates that have incorporated samples from the Chatham Islands (e.g., *Paphies subtriangulata*, Smith et al. 1989; *H. virginea* Clarke 2001; *Cellana strigilis*, Goldstien 2006a). Differentiation between *H. iris* from the mainland and *H. iris* from the Chatham Islands probably resulted from isolation by distance as supported with the Mantel tests.

Although migration rates were not calculated, the Chatham Islands and the North and South Island sites shared haplotypes. The presences of shared haplotypes and weak genetic structure suggested limited gene flow between the Chatham Islands and the mainland; however, shared haplotypes could also result from homoplasy or incomplete lineage sorting. Passive transport of *H. iris* larvae via the Southland Current to the Chatham Islands seems unlikely. Based on a Southland Current speed of 0.2 ms^{-1} (Hadfield et al. 2007), passive transport of abalone larvae would take around 52 days to traverse 900 km. Even with an extended larval time of 22 days (Roberts and Lapworth 2001), *H. iris* larvae would still be incapable of travelling to Chatham Islands by passive transport alone. Alternatively, migration of *H. iris* between the North and South Islands and the Chatham Islands may occur by rafting on drifting macroalgae or active transport by humans.

Rafting on drifting macroalgae has been hypothesized as a means of long distance dispersal for many marine invertebrates (Highsmith 1985; Holmquist 1994; Worcester 1994; Watts et al. 1998; Castilla and Guinez 2000; Sponer and Roy 2002; Aliani and Molcard 2003; Grantham et al. 2003; Thiel and Gutow 2005). Very few effective migrants per generation are required to counter the effect of genetic drift and prevent population differentiation (Hartl and Clark 1997), thus even sporadic dispersal via drifting macroalgae would be effective at homogenizing populations. Some species even have physiological and behavioral attributes that could facilitate rafting. For example, an obligate rafter, *Idotea metallica*, utilizes food more efficiently and accumulates more lipids than a facultative rafter, *I. baltica*, enabling *I. metallica* to cope better with the potentially low food availability of their rafting lifestyle (Gutow et al. 2006). For another example, McCormick et al. (2008) identified two behaviors that would promote rafting on macroalgae in the white abalone, *H. sorenseni*. In the presence of drifting macroalgae, many juveniles and young adults would ‘stand’ and then ‘climb’ onto stands of drifting kelp. McCormick et al. (2008) suggested that such behavior facilitates rafting and explains the presence of *H. sorenseni* on isolated rock outcrops and beyond the

range of larval dispersal. Potentially, *H. iris*, often found in beds of macroalgae, could exchange migrants between the North and South Islands and the Chatham Islands by rafting.

Human mediated transport of abalone might also be responsible for shared haplotypes. As a commercially and culturally important commodity, abalone are transferred around New Zealand. Present day movement of abalone around New Zealand might include releasing of poached individuals, discarding of individuals at processing plants, the escape of captive individuals from aquaculture farms, and the deliberate release of juveniles to enhance stocks. For example, movement of *P. canaliculus* around New Zealand to seed mussel farms has led to the introgression of genes from northern populations into southern populations (Apte et al. 2003). The extent of abalone trafficking is unknown; although, a few instances of attempts at stock enhancement have been documented (Schiel 1993; Roberts et al. 2007). Possibly, anthropogenic movement of *H. iris* has resulted in low levels of gene flow between the Chatham Islands and mainland New Zealand.

Instead of migration, homoplasy and incomplete lineage sorting may be distorting the picture of gene flow between the Chatham Islands and mainland New Zealand. Shared mitochondrial haplotypes may have arisen independently in individuals at multiple sites; however, the chance of this occurring for three different haplotypes (8, 17, and 18), two to nine mutations apart, seems very unlikely. On the other hand, the haplotypes may be ancestral. Present day gene flow may be absent, but may have persisted in the past. Thus, the pattern observed may simply reflect that not enough time has passed for the Chatham Islands and mainland New Zealand to acquire reciprocal monophyly.

North Island vs. South Island

Significant structure existed among mainland samples after removal of the Chatham Islands sample. AMOVAs rejected the hypothesis of homogeneity around the Cook Strait region, but unlike other New Zealand invertebrates *S. pelliserpentis* (Veale 2007), *C. ornata* (Goldstien et al. 2006b), *C. radians* (Goldstien et al. 2006b), *P. regularis* (Ayers and Waters 2005; Waters and Roy 2004), and *P. canaliculus* (Apte and Gardner 2002; Star et al. 2003), the structuring did not correspond to known regions of upwelling (Figure 2.1). Instead, the largest Φ_{CT} value occurred when samples were partitioned across Cook Strait narrows. However, none of structures proposed in Table 2.2 might best explain the genetic structure of *H. iris.*

Although MDS had a high stress value and the neighbor joining method can be spurious, both clustering techniques hint at alternative structures. No clear groups were

produced with MDS; however, OLB was positioned closer to the South Island samples than the North Island samples. Clustering based on the neighbor joining method grouped the TCL sample in Cook Strait with North Island samples, the PHD sample in Cook Strait with South Island samples, and the three samples on the east coast of the North Island (EAI, OLB, and MAT) with the South Island samples. Running AMOVAs among mainland samples (excluding OCH) with either of these structures produced larger Φ_{CT} values than any a priori structure listed in Table 2.2. Of the two structures suggested by the clustering techniques, the neighbor joining structure (EAI, MAT, OLB with the South Island vs. TCL with the North Island) produced a larger Φ_{CT} (0.04722, p = 0.00) than the MDS structure (Φ_{CT} = 0.04350, p = 0.00).

The similarity between TCL and the North Island may be related to the physical distance between TCL and WLG and the large amount of mixing in the Cook Strait region (Harris 1990). The grouping of three North Island samples with the South Island samples may be related to the presence of the Wairarapa Coastal Current (WCC in Figure 1.4, Chiswell 2000). The Southland Current mixes with the D'Urville Current east of Cook Strait, and this mixed water forms the WCC, which flows north along the east coast of the North Island. Modeling particle movement in and around Tolaga Bay (OLB) indicated a northward flow of particles, but this was probably due to wind driven currents (Stephens et al. 2006). Such northward movements may be responsible for the sharing of haplotypes 17 and 18, which are more predominant in the South Island and Chatham Islands, with the east coast of the North Island.

Alternatively, the split may represent mutational and/or demographic differences between northern and southern samples. The north-south split is reflected in the abundance of rare haplotypes and high haplotype diversity in the northern samples. Identifying possible causes for such patterns may result in alternative hypotheses for future testing.

First, higher haplotype diversity in samples from the North Island might result from increased mutation rates due to warmer water. Mutation rates differ between species (Bromham and Penny 2003), and differences may result from varying metabolic rates, which are affected by temperature (Martin and Palumbi 1993; Bleiweiss 1998; Gillooly et al. 2005; Estabrook et al. 2007; Gillooly et al. 2007). Organisms with higher metabolic rates are expected to have higher mutation rates. However, this hypothesis is contentious (Held 2001; Lanfear et al. 2007) and may be applicable to only a few species and a few genes (Lanfear et al. 2007). The studies that test this hypothesis concentrated on interspecific comparisons with phylogenies that span millions of years. During such time spans species evolve and

environments change such that rates of evolution represent an average. At any specific point in time, this hypothesis may be applicable to a species with demes distributed over a broad range of temperatures. For *H. iris*, potentially the northern samples inhabiting warmer water (Figure 2.11) have an increased mutation rate resulting in higher haplotype diversity for the northern samples.

Second, the reduced haplotype diversity in southern populations and the Chatham Islands might result from varying fishing pressures. Intense fishing is suspected to be a selective agent changing the genetic makeup of a population (Allendorf et al. 2008). For instance, Hauser et al. (2002) genotyped six microsatellite loci from samples of New Zealand snapper (*Pagrus auratus*) collected from 1950–1986 and 1998. Their study documented the loss of genetic diversity (mean heterozygosity and mean number of alleles) in an exploited New Zealand stock. *H. iris* are a heavily exploited species controlled by a quota system. New Zealand's coasts are divided into nine management areas with each area allotted a Total Allowable Commercial Catch (TACC, Figure 2.12). The larger TACC for the South Island and Chatham Islands may act to bottleneck *H. iris* in these areas. As a result, rare alleles are removed, haplotype diversities are reduced. Interestingly, the South Island samples were not significantly different from the Chatham Island samples (AMOVA: $\Phi_{CT} = 0.11420$, $p = 0.071$), but the South Island and Chatham Island samples combined were significantly different from the North Island samples (AMOVA: $\Phi_{CT} = 0.03622$, $p = 0.000$).

In contrast to the clustering techniques and the AMOVAs, SAMOVAs did not identify structure between northern and southern samples consistent with those proposed in Table 2.2. Although a simulation study showed that SAMOVAs always identify maximally differentiated groups, they did not always identify the correct group (Dupanloup et al. 2002). In fact with a single locus and assuming a stepping stone model, SAMOVAs identified the correct group 92.4–96.0% of the time *only* when migration between groups was low and the migration within groups migrations were high (1000 times larger than that between groups. In all other cases, SAMOVA performed much worse and identified the correct group 2.3–57.2% of the time. The low Φ_{ST} and Φ_{CT} values combined with many shared haplotypes among sampling locations suggested that migration rates were high making SAMOVAs less reliable at identifying the correct group.

Ideally, SAMOVAs should have grouped geographically adjacent samples (Dupanloup et al. 2002), yet it grouped non-adjacent samples. For example, it grouped DBL and GLN but did not include the geographically intermediate SPB sample. This in part may have been due to the use of coordinate distances rather than coastal distances or it may have

been due to the SAMOVA algorithm itself. Dupanloup et al. (2002) noted that SAMOVAs were capable of defining groups in which not all members are geographically adjacent; however, they offered no explanation of this behavior.

SAMOVAs tended to isolate single samples, possibly due to a lot of within sample variation or stochastic variability in DNA sequences. However if SAMOVAs identified the correct groups, then a lot of the sampling locations would be considered isolated (e.g., IHM, DBL, GLN, EAI, or CBL) possibly due to local recruitment. Another potential group would include TIM and MTB, consistent with them being differentiated from most other samples (Table 2.8). Isolation of these samples from the rest of the South Island might be related to coastal circulation, whereas the offshore Southland Current might play a bigger role in the transport of abalone between the remaining South Island. Further exploration with population genetic or oceanographic studies could elaborate the extent and potential causes of these alternative groupings.

The different techniques (MDS, neighbor joining clustering, AMOVAs, SAMOVAs), all identified slightly different structures among mainland samples. This combined with the low level of population divergence (Φ_{ST} = 0.03815) emphasized that *H. iris* did not have a clear-cut genetic structure, as opposed to the level and pattern of genetic structures identified in *C. ornata* (Φ_{ST} = 0.829, Goldstien 2005, 2006b) and *S. pelliserpentis* (Φ_{ST} = 0.45, Veale 2007). Instead, the pattern of genetic structure for *H. iris*, resembled those species with more intermediate levels of population genetic structure (e.g., *P. regularis* Φ_{ST} = 0.072, Waters and Roy 2004; *P. canaliculus* Φ_{ST} = 0.162, Apte and Gardner 2002; *C. radians* Φ_{ST} = 0.142, Goldstien 2006b).

MDS analyses for *P. regularis* (Waters and Roy 2004; Ayers and Waters 2005) and *C. radians* (Goldstien 2005) presented samples as a continuous spectrum (similar to *H. iris*) rather than distinct clusters (no cluster analysis was presented for *P. canaliculus*). The lack of concordance in pattern and level of population differentiation among New Zealand coastal invertebrates suggested that the effects of potential barriers to gene flow in the Cook Strait are species specific. As previously discussed, larval behavior and life history characteristics (Hedgecock 1986; Bohonak 1999; Goldstien 2005; Veale 2007) and/or demographic differences may limit the influence of the Cook Strait region on population genetic structure.

Sample sizes

The number of individuals sampled per location was small given that only a single locus was used and a low level of population differentiation was observed. For comparison

when 16 loci are used, sample sizes of 20 individuals would be reasonable when $\Phi_{ST} = 0.05$, while sample sizes of 100 individuals would be reasonable when $\Phi_{ST} = 0.01$ (Kalinowski 2005). Reducing either the number of loci used and/or the sample size per location (as in this study) increases the coefficient of variation for genetic distance making estimates less precise. To increase precision, future study could either sample more individuals, pool samples based on a priori population relationships, use more variable markers, and/or incorporate more molecular markers (Kalinowski 2002a, Kalinowski 2005).

Future work

H. iris around New Zealand have intraspecific structure. As this project is ongoing, several more location samples have been collected on the west coast of the South Island and the southeast coast of the North Island. These samples will be sequenced for the same regions and incorporated into the analysis. Based on the haplotype frequencies found here, the South Island sample should have smaller haplotype diversity than the North Island sample. Furthermore because the North Island sample is from the east coast of the North Island, it is expected to share some of the South Island haplotypes like haplotypes 17 and 18. Although, mitochondrial DNA is popular marker for phylogeographic studies, it does not necessarily reflect variation in the nuclear genome. Nuclear markers, e.g. microsatellites, are currently being developed for comparison and the progress in this research is presented in Chapters 3 and 4.

.

3. Isolation and characterization of *Haliotis iris* microsatellites

ABSTRACT

Microsatellites are useful in population genetic studies because of their high levels of polymorphisms. Over 380 microsatellites have been isolated for a range of abalone species; however, prior to this study, no microsatellites existed for *Haliotis iris*. Since the development of microsatellites de novo takes time and money, primer pairs isolated for 11 microsatellite loci in *H. midae* and 8 microsatellite loci in *H. rubra* were trialed for amplification of *H. iris* DNA. Unfortunately, only two primer pairs consistently amplified *H. iris* DNA, of which one amplified a monomorphic repeat and the other amplified a fragment containing no microsatellite repeat. As a result, 13 microsatellite loci were isolated from *H. iris* DNA. Of these, eight loci were deemed unsuitable for further studies because primers amplified more than two alleles in a single individual, alleles differed greatly in size, and/or the allelic electropherogram patterns had different shapes. Three loci were optimized to produce clean bands and were screened across 25 samples of *H. iris* used in Chapter 2. These loci were very polymorphic (number of alleles ranged from 23 to 84) and should be useful for further population genetic studies.

INTRODUCTION

Microsatellites are frequently employed in population genetic studies due to their fast mutation rate and abundance. Microsatellites are short (1–6 bp) tandemly repeated DNA sequences. The string of short tandem repeats result in fast mutation rates due to errors in replication such as unequal crossing over, gene conversion, and replication slippage. Microsatellite mutation rates are locus- and species-specific and vary from 10^{-6} to 10^{-2} per locus per generation (reviewed in Schlotterer 2000; Ellegren 2004; Nikitina and Nazarenko 2004; Oliveira et al. 2006). For comparison, human microsatellites mutate at rates varying from 10^{-6} to 10^{-4} per locus per gamete per generation (Nikitina and Nazarenko 2004 and references therein), while human single copy nuclear sequences mutate at an approximate rate of 10^{-9} per nucleotide per generation (Nikitina and Nazarenko 2004), and primate mitochondrial DNA mutate at an approximate rate of 10^{-8} per nucleotide per generation (Brown et al. 1979). The high mutation rates mean that, in general, microsatellites will be more polymorphic than nuclear sequence data, and markers with high polymorphism are desirable for more precise estimates of population genetic structure (Kalinowski 2002a; Kalinowski 2005). Furthermore, microsatellites are also abundant in the genome (Li et al. 2002b; Tóth et al. 2000). For example on average in humans, one microsatellite occurs in every 2000 bp (Lander et al. 2001 as cited in Nikitina and Nazarenko 2004). As a result, large

numbers of independent loci can be isolated for genetic analysis (e.g., Baranski et al. 2006; Sekino et al. 2006).

Microsatellites are codominant markers that are analyzed with respect to the frequency of different sized alleles. The relationship among microsatellites cannot be readily examined directly, but can be inferred using explicit evolutionary models. For instance, the two predominant models in population genetic studies are the infinite allele model (IAM, Kimura and Crow 1964) and the stepwise mutation model (SMM, Kimura and Ohta 1978). The IAM assumes that each mutation creates a new allele, while the SMM assumes that each mutation results in a stepwise change from the last allelic state, in the case of microsatellites this is assumed to be the gain or loss of a repeat. Under the SMM, similar sized alleles are more closely related than alleles of greatly different sizes, whereas under the IAM alleles of similar size are expected to be no more closely related than alleles that differ markedly in size. Related to the IAM and the SMM are numerous alternative models like the K-allele model (KA, Crow and Kimura 1970) and the two phase model (TP, Di Rienzo et al. 1994). The KA model limits the number of alleles (K) for a given locus and includes a probability for mutating into another allele, while the IAM model is the KA model with $K = \infty$. The TP model expands the SMM model by assuming the gain and loss of a single repeat is common with an occasional gain or loss of a large number of repeats (Di Rienzo et al. 1994).

Unfortunately, the details of microsatellite evolution are still being uncovered leaving the accuracy of various models debatable and the proposition of new models inevitable (reviewed in Li et al. 2002b; Ellegren 2004; Nikitina and Nazarenko 2004; Oliveira et al. 2006). Another drawback of microsatellites is homoplasy—when the same allele arises in two individuals independently. Homoplasy is potentially problematic in markers like microsatellites that have high mutation rate and allele size constraints. Nonetheless, microsatellites are frequently used to examine intraspecific and shallow interspecific differentiation.

Haliotis and conserved microsatellites

The commercial importance of abalone species and the desire to identify and conserve endangered populations have driven the development of microsatellites as tools for studying population genetic structure (e.g., Evans et al. 2000; Miller et al. 2001; Cruz et al. 2005), examining the effects of stock enhancement using hatchery bred abalone (e.g., Selvamani et al. 2001; Bester et al. 2004; Gutiérrez-Gonzalez and Perez-Enriquez 2005), and mapping in selective breeding projects (e.g., Selvamani et al. 2000; Li et al. 2002b; Baranski et al. 2006).

As a result, over 380 microsatellites have been developed for a variety of commercially important abalone species around the world (Appendix 5). Instead of developing microsatellites de novo for each species of abalone, time and money could be saved by adopting preexisting microsatellite primers for new abalone species. Cross species amplification of microsatellites will depend on the conservation of the primer sites: lack of amplification only indicates that the primer sites have been lost and not necessarily the microsatellite.

A number of studies have reported cross-amplification of abalone microsatellite primers (Appendix 5). Arguably, the most successful cross-amplification of abalone microsatellites involved microsatellite primers for *H. kamtschatkana* (Miller et al. 2001). *H. kamtschatkana* microsatellites have been successfully used to study genetic variation in *H. rufescens* (Gruenthal et al. 2007), *H. sorenseni* (Gruenthal and Burton 2005), *H. corrugata* (Díaz-Viloria et al. 2008), and *H. fulgens* (Gutiérrez-Gonzalez and Perez-Enriquez 2005; Gutiérrez-Gonzalez et al. 2007; Díaz-Viloria et al. 2008). The success of the *H. kamtschatkana* primers may be a result of the close evolutionary relationship among these species (Figure 1.2).

Although cross-amplification studies have indicated successful amplifications occur more in sister taxa (Huang and Hanna 1998; Evans et al. 2001), evolutionary distance between species is not always a predictor of successful cross-amplification. For instance, more *H. rubra* microsatellites were suitable for studying genetic variation in *H. midae* than the more closely related *H. laevigata* (Evans et al. 2001). Although two more primer pairs amplified *H. laevigata* DNA, only 25% of the successful primer pairs could be optimized to produce suitable microsatellites for *H. laevigata*, while 60% of the successful primer pairs could be optimized to produce suitable microsatellite for *H. midae* (Evans et al. 2001). These findings also highlighted that simply showing cross-species amplification does not imply the microsatellite exists or will be useful in other species.

H. iris and microsatellites

The aims of this chapter were 1) to develop a set of microsatellite markers for *H. iris*, and 2) use these markers to genotype the same samples used in Chapter 2. Further examination of *H. iris* genetic structure using microsatellites occurs in Chapter 4. As corroborated by Selkoe and Toonen (2006), this study first searched for *H. iris* microsatellites by examining already developed haliotid microsatellites (Appendix 5). No previous microsatellites existed for *H. iris*, but Evans et al. (2001) found eight *H. rubra* microsatellite

primer pairs that amplified *H. iris* DNA. In seven of the eight cases where microsatellite primer pairs amplified *H. iris* DNA, the primers also amplified *H. midae* DNA. In addition, Bester et al. (2004) published another 11 microsatellite primer pairs for *H. midae*. As a result, cross-amplification of *H. iris* DNA with *H. rubra* and *H. midae* microsatellite primer pairs were tested. In addition to cross-species amplification, the University of Canterbury and OceaNZ independently pursued the development of *H. iris* microsatellites. This chapter presents the characterization of only the University of Canterbury microsatellites.

MATERIALS AND METHODS

Cross-species amplification

Since microsatellite development takes time and money, microsatellite primers from *H. rubra* (Evans et al. 2001) and *H. midae* (Bester et al. 2004) were initially tested on samples of *H. iris* (Table 3.1). Initial trial amplifications included five *H. iris* DNA samples, two *H. midae* samples (as positive controls), and a negative control. DNA samples were extracted using Qiagen's DNEasy® Blood & Tissue Kit. PCR reactions had a total volume of 15 μL and contained 1–20 μL of genomic DNA, 200 μM dNTPs, 0.33 μM of each primer, 2.5 mM $MgCl_2$, 1X NH_4 Reaction Buffer (160mM $(NH_4)_2SO_4$, 670 mM Tris-HCl (pH 8.8 at 25 °C), and 0.1% Tween-20), and 0.6 units BIOTAQ™ (Bioline USA, Inc.). Thermal cycling parameters were an initial denaturation at 95 °C for 12 minutes, 10 cycles of 94 °C/15 s, 45 °C/15 s, 72 °C/15 s, 25 cycles of 89 °C/15 s, 45 °C/15 s, 72 °C/15 s, and a final extension of 72 °C for 10 minutes. Follow-up PCRs included varying the annealing temperatures from 45–60 °C, the magnesium concentrations from 1.0–5.0 mM, and the amount of DNA. Amplifications were checked on 2% agarose gels against HyperLadder V™ (Bioline USA, Inc.) and were visualized under UV following staining with ethidium bromide (0.5 μg/mL). Loci that produced clean amplification, HmD59 and HmSP5, were sequenced following protocols described in Chapter 2.

Promising amplifications were size-fractionated on an ABI3100 Genetic Analyzer (Applied Biosystems, Inc.) at the University of Canterbury. Genomic DNA was then amplified with optimized conditions and an additional ingredient, 0.5 μM of ChromaTide® Rhodamine Green™-5-dUTP (Invitrogen™). Normally, microsatellites are amplified with fluorescently labeled primers (Becher and Griffiths 1997). To avoid purchasing fluorescently labeled primers, fluorescently labeled dNTPs (dUTPs) were used. Microsatellites amplified with dUTPs will produce a unique spiky shape on an electropherogram (Figure 3.1). Since the

dUTPs are incorporated in each strand and the strands have different base compositions requiring different amounts of dUTPs, the strands run slightly differently resulting in a spiky shape. To reduce the noise of extra dUTPs, amplification products were precipitated with ethanol and eluted in molecular grade water. Precipitations were checked on 2% agarose gels against HyperLadder V™ (Bioline USA, Inc.) and visualized with ethidium bromide 0.5 μg/mL. For screening on an ABI3100, 1 μL of cleaned product was added to 14 μL of HiDi™ Formamide (Applied Biosystems, Inc.) and 0.3 μL of GeneScan™-500 LIZ® Size Standard (Applied Biosystems, Inc.). Mixtures were heated for 2 minutes at 95 °C and then chilled on ice for 2 minutes. Electropherograms were viewed using GeneMarker v1.6 Demo (SoftGenetics®).

Simultaneous to this study, Baranski et al. (2006) were developing a panel of microsatellite primers for genetic mapping in *H. rubra.* Twelve *H. iris* DNA samples were sent to Dr. Helen McPartlan's laboratory (Department of Primary Industries, Australia). Shannon Loughnan graciously screened 18 primer pairs developed for *H. rubra* from a set of 35 primer pairs that produce high quality polymorphic genotypes in *H. rubra* and *H. laevigata*. Unfortunately, none of the 18 primer pairs amplified *H. iris* DNA, and no more primers were pursued.

Microsatellite development and screening

Unsuccessful cross-species amplification left no alternative other than developing *H. iris* specific microsatellites. Microsatellite development protocols are readily available in the literature (reviewed in Zane et al. 2002), and the development of such markers is routine. However, the majority of microsatellite isolation techniques require cloning. Under the Hazardous Substances and New Organisms (HSNO) Act of 1996 such modification constitutes the formation of a new organism and is prohibited without a permit. Additionally, work on native species is particularly controversial and requires consultation with iwi prior to seeking regulatory approval. The entire process can be very time consuming and costly, thus given limited time and funds available for this project, microsatellites were developed by ATG genetics Inc. (Vancouver, Canada) from ten *H. iris* DNA samples, representing ten different sampling locations: DSD, DUI (equivalent to PHD), GOB, MAT, MTB, NPT, OLB, TAB (equivalent to DBL), TPI (equivalent to OCH), and WHN. From these samples, the ATG genetics' Starter Kit used GATA enrichment to isolate and develop primer pairs for 13 unique polymorphic microsatellite loci (Table 3.2 and Appendix 6).

The primer pairs identified by ATG genetics, Inc. were tested and optimized across three *H. iris* DNA samples and a positive control at the University of Canterbury. Thermal cycling parameters consisted of an initial denaturation at 95 °C for 12 minutes, 10 cycles of 94 °C/15 s, 50 °C/15 s, 72 °C/15 s, 25 cycles of 89 °C/15 s, 50 °C/15 s, 72 °C/15 s, and a final extension of 72 °C for 20 minutes. Annealing temperatures of 50 °C were chosen because they were 5 °C lower than the annealing temperatures supplied by ATG genetics, Inc.. PCR reactions had a total volume of 15 μL and contained 1–20 μL of genomic DNA, 200 μM dNTPs, 0.33 μM of each primer, 2.5 mM $MgCl_2$, 1X NH_4 Reaction Buffer (160 mM $(NH_4)_2SO_4$, 670 mM Tris-Hcl (pH 8.8 at 25 °C), and 0.1% Tween-20), and 0.6 units BIOTAQ™ (Bioline USA, Inc.). Amplifications were checked on 2% agarose gels against HyperLadder V™ (Bioline USA, Inc) and were visualized under UV following staining with ethidium bromide (0.5 μg/mL).

Microsatellite primer pairs were further optimized across more samples to increase the consistency of amplification and reduce the number of spurious or systematic extra bands. In some cases, new primers (Table 3.2) were designed with Primer3 (Rozen and Skaletsky 2000). Optimization techniques included reducing the overall number of cycles, varying annealing temperatures from 45–65 °C, and changing the concentration of $MgCl_2$ from 1.5–2.5 mM. To confirm the appropriate band was optimized, microsatellites were amplified with fluorescently labeled dUTPs and size-fractionated on an ABI3100 Genetic Analyzer (Applied Biosystems Inc.) at the University of Canterbury, as previously described. Microsatellites were evident, but fragments were difficult to score. As an alternative, seven forward primers of the more promising primer pairs (Table 3.2) were ordered with M13 sequence tags. PCR reactions were modified to incorporate 0.09 μM of M13 tagged forward microsatellite primer, 0.3 μM of untagged reverse microsatellite primer, and 0.3 μM fluorescent tagged M13 forward primer (Schuelke 2000). However, microsatellites were still difficult to score. Primers were reordered with fluorescent tags (Table 3.2). Fluorescently labeled primers were optimized similar to above and screened across 7–18 individuals (Table 3.3).

Ultimately only three loci (AB14, AB21, and AB31) were suitable for population genetic analysis. The same 477 individuals used in Chapter 2 (Figure 2.3) were genotyped at these three loci. Microsatellite PCRs had a total volume of 15 μL and contained around 1–20 ng of DNA and followed the optimized PCR conditions (Table 3.4). PCRs were performed separately for each locus, checked on 2% agarose gels against HyperLadder V™ (Bioline USA, Inc), and were visualized under UV following staining with ethidium bromide (0.5

μg/mL). Depending on amplification intensity 1–3 μL of PCR products were mixed with 9–12 μL HiDi™ Formamide (Applied Biosystems, Inc.) and 0.3 μL of GeneScan™-500 LIZ® Size Standard (Applied Biosystems, Inc.). The slurries were denatured for 3 minutes at 95 °C and held on ice for 2 minutes. Amplified fragments were size-fractionated on an ABI3100 Genetic Analyzer (Applied Biosystems Inc.) at the University of Canterbury. Electropherograms were analyzed using GeneMarker v1.6 Demo (SoftGenetics®).

Allele scoring

Most individuals had clear and easily distinguished alleles (Figure 3.2). For all loci, the size of the right most peak was recorded to avoid discrepancies due to the incomplete addition of a 3' adenosine (Smith et al. 1995). Although microsatellites with tetranucleotide repeats are supposedly easier to score than microsatellites with dinucleotide repeats, the continuum of alleles present for these loci, especially for AB21, made scoring difficult. Binning possibilities were explored using bin sizes of one, two, and four repeats in FlexiBin (Amos et al. 2007). Using a combination of FlexiBin outputs, the maximum difference in allele sizes for individuals that were genotyped more than once, and the electropherograms, alleles were labeled with a three digit number reflecting the total fragment length. In general, the binning resembled FlexiBin results for a bin size of a one base pair repeat with a few modifications.

Loci in individuals that did not amplify, contained messy electropherograms, or more than two alleles per locus per individual were scored as missing data. Given several reactions that did not amplify and the frequently reported null alleles in other mollusk microsatellite studies (e.g., Hedgecock et al. 2004; Baranski et al. 2006), null alleles were calculated according to Brookfield (1996) implemented in MICRO-CHECKER 2.2.3 (Van Oosterhout et al. 2004). MICRO-CHECKER uses randomizations to test for departures from Hardy-Weinberg equilibrium and examines patterns across genotypes to distinguish between null alleles, large allele dropout, scoring errors due to stuttering, and deviations from panmixia.

Genotyping error rate

Null alleles, low quantity or quality of DNA, PCR artifacts, allele stuttering, and typos are notorious sources of error when scoring genotypes (detailed in Bonin et al. 2004; Hoffman and Amos 2005; Pompanon et al. 2005; DeWoody et al. 2006). To assess error rate, two groups of eight from the six extraction plates were re-amplified for each locus. Failed reactions were not repeated. Electropherograms and allele sizes were compared between

original amplifications and re-genotyped amplifications. Only instances of both clear original and clear re-genotyped amplifications (e.g., Fig 3.2) were used to assess the number of original amplifications that were correct and incorrect. Error rates were calculated per reaction and per allele as in Hoffman and Amos (2005) and Pompanon et al. (2005).

Locus characterizations

To characterize the loci, all samples were grouped into one population. Observed and unbiased expected heterozygosities were calculated in GeneAlEx 6.1 (Peakall and Smouse 2006). Deviations from Hardy-Weinberg equilibrium were calculated using an exact test based on a Metropolis-Hastings Markov chain algorithm (Guo and Thompson 1992) and the default parameters (dememorization length = 10,000, batch length = 100, iterations per batch = 5,000) in GENEPOP 4.0 (Rousset 2008). Gametic disequilibrium between loci was also tested using a Markov chain algorithm (Raymond & Rousset 1995) in GENEPOP 4.0 (Rousset 2008).

RESULTS

Cross-species amplification

All of the eight *H. rubra* and eleven *H. midae* microsatellite primers produced a visible PCR product; however, none were found to amplify a clean and variable product from *H. iris* DNA (Table 3.1). Attempts to optimize all *H. rubra* and nine of the *H. midae* primers failed: PCRs were inconsistent, not enough amplicons were produced, or too many different sized amplicons were produced. Primers for two *H. midae* loci, HmD59 and HmSP5, consistently amplified a single band on agarose gels. When size-fractionated on the AB3100, the HMD59 amplicon was a single peak was of 220 bp, while HmSP5 amplicon was a single peak of 412 bp. Both of these peaks were larger than the expected allele size ranges (106–150 bp and 189–215 bp, respectively) for *H. midae* (Bester et al. 2004). To examine whether the amplifications contained a microsatellite, they were sequenced. The HmD59 locus contained an $(AG)_3$ sequence, while HmSP5 locus contained no sequence repeated more than twice. In *H. midae*, both loci contained an $(AC)_N$ sequence (Bester et al. 2004). Furthermore, neither HmD59 or HmSP5 *H. iris* sequences resembled the *H. midae* sequences available on GenBank (ACCN: AY303336 and AY303344, respectively).

Screening H. iris loci

Thirteen polymorphic loci were developed by ATG genetics, Inc. (Table 3.2), which were variously screened for polymorphism using direct incorporation of fluorescently labeled dUTPs and/or fluorescently labeled primers. A fluorescently labeled M13 primer method was trialed on seven loci (Schuelke 2000). This method did not produce easily scoreable alleles. Many PCR artifacts (possibly due to the M13 tags) were present making it difficult to achieve adequate intensities to reproducibly score alleles. As a result, primers for these seven loci plus primers for an additional five loci, which showed successful amplification on agarose, were ordered with fluorescent tags (Table 3.2). Of these 11, eight were deemed unsuitable, and three (AB14, AB21, and AB31) were used in subsequent population level analyses (Table 3.3). The key problems identified with the eight unsuitable loci included the presence of more than two alleles per individual, large size ranges between alleles, large alleles beyond the sizing range of the size standard, and different shaped alleles and stutter bands in the electropherograms (Table 3.3, Figure 3.3).

Genotyping error rate

To assess potential problems with the genotyping (i.e., null alleles and/or human error), genotyping error rates were calculated for each locus (Table 3.5). Locus AB14 had the lowest genotyping error rate per reaction at 3.6 %, while AB31 had the lowest error rate per allele at 3.4 % (Table 3.5). For AB21, four alleles were mis-entered, and two alleles (one reaction) were different from the alleles produced in the re-genotyped amplification. For AB31, four alleles (three reactions) were mis-entered, and two alleles (one reaction) were different from the alleles produced in the re-genotyped amplification.
Out of the 59 complete genotypes, nine had an error in at least one allele, giving an error rate per multilocus genotype of 15.3 %.

Characterization of AB14, AB21, and AB31

To assess levels of polymorphism, the 477 *H. iris* from 25 locations around New Zealand (Figure 2.4) were genotyped. Out of these samples, 18 individuals failed to amplify with the AB14 primers, 28 individuals failed to amplify with the AB21 primers, and 17 individuals failed to amplify with the AB31 primers. Amplifications failed at two loci for six individuals and at three loci for another five individuals. All loci were polymorphic in each sampling locality. The total number of alleles per microsatellite locus ranged from 23–84, and the average expected heterozygosities ranged from 0.825–0.973 per locus (Table 3.6). All loci

had a significant deficiency of heterozygotes. No pairs of loci were in gametic disequilibrium. Null alleles were estimated to be present in all loci (Table 3.6).

DISCUSSION

After laborious testing of 13 potential microsatellites, three microsatellite markers, AB14, AB21, and AB31, were successfully optimized and used to genotype 477 *H. iris* from 25 locations around New Zealand (explored more in Chapter 4). Although only three microsatellites were suitable, they were extremely polymorphic, and few highly polymorphic loci may provide just as precise estimates of genetic distance as more loci with few alleles (Kalinowski 2002a). According to Kalinowski 2002a, the number of independent alleles AB14, AB21, and AB31 had 22, 83, and 41 independent alleles for a total number of independent alleles of 146. When population divergence is small, these loci should produce a similar amount of precision for genetic distance estimates as 73 loci each with two independent alleles.

Cross-species amplification

Attempts at using microsatellite primers developed for *H. midae* and *H. rubra* failed. Fragments screened on the ABI3100 had too much noise and no characteristic microsatellite pattern. Successful amplifications identified on agarose did not imply microsatellite conservation. As expected, all the *H. rubra* primers amplified *H. iris* DNA (Evans et al. 2001), but none could be optimized as a suitable marker. The *H. midae* microsatellite primers (Bester et al. 2004) have never been trialed on another species until now. As with the *H. rubra* primers, *H. midae* primers amplified *H. iris* DNA, but none could be optimized as a suitable marker. The loci with the most potential, HMD59 and HmSP5, proved to be either monomorphic (HmD59) or did not contain a repetitive motif (HmSP5).

The lack of cross-species amplification and the need to develop microsatellites de novo are drawbacks to using microsatellites in many species. In haliotids, successful cross-species amplifications have been attributed to evolutionary distance between taxa (Huang and Hanna 1998; Evans et al. 2001). Accordingly, the unresolved position and the potential lack of a closely related sister taxa for *H. iris* (Figure 1.2) suggest cross-species amplification of microsatellites between *H. iris* and other haliotids would be rare. However, evolutionary distance may not be the sole factor in determining successful cross-species amplification. Barbará et al. (2007) suggested that lack of cross-species amplification may result from the accumulation of mutations. Their meta-analysis of 611 marker transfer experiments (from 64

different primer notes) found that mating system, generation time, and genome size, as well as taxonomic grouping, were indicative of successful marker transfers: successful transfers will be less likely in semelparous, short-lived individuals with a small genome size. Likewise, marine broadcast spawners, like abalone, may accumulate mutations quickly because of their high fecundity, large variance in reproductive success, and potentially small N_e (Hedgecock 1994; Hedrick 2005). Due to the stochastic survival of offspring carrying mutations coupled with a small N_e, new mutations may be rapidly fixed in abalone and decrease the likelihood of cross-species amplification.

H. iris microsatellites

Although 13 microsatellite loci were isolated de novo, only three loci were suitable for further population genetic analysis, while the remaining loci could not be scored consistently due to the presence of more than two alleles in an individual, large allele size ranges, and different shaped alleles (Table 3.2). First, fluorescently tagged primers for two loci produced three or four alleles in single individuals. These extra alleles could have resulted from contamination; however, equivalent intensities (when adjusted for allele size) of the alleles on the electropherograms (Figure 3.3A) and the production of at most two alleles for other loci amplified from the same DNA suggested the samples were not contaminated. The extra alleles probably resulted from a duplicated locus. Baranski et al. (2006) isolated 125 loci for *H. rubra* and found that one individual had three alleles for four loci and progeny in a known cross had more than two alleles for two other loci. Noting very similar flanking sequences among their loci and in comparison to abalone microsatellites and genes listed on GenBank, Baranski et al. (2006) attributed the extra alleles to frequent locus duplications or associations with mobile elements.

Although the ploidy level of *H. iris* is unknown, 15 abalone species are known to be diploid (2n = 28–36, Gallardo-Escárate et al. 2004; Hernández-Ibarra et al. 2004 and references therein); yet, peculiarities in abalone karyology suggest that locus duplication might be common. In comparison with other vetigastropods and basal gastropods in which 2n = 18–20 (Patterson 1967 and Haszprunar 1988, as cited in Geiger 1999), abalone have more chromosomes. As a result, Geiger and Groves (1999) proposed that abalone evolution has been marked by a progression of increasing ploidy level.

Although the frequency of triploids, tetraploids, and aneuploids in the wild are unknown, abalone ploidy levels like those of other molluscs (e.g., *Crassostrea gigas*, *Placopecten magellanicus*, and *Mytilus edulis*, Desrosiers 1993) can be manipulated in

laboratories with reagents like caffeine (Okumura et al. 2001), 6-dimethylaminopurine (6-DMAP, Liu et al. 2004), or cytochalasin B (Yang et al. 1998a; Maldonado et al. 2001; Liu et al. 2004; Li et al. 2007b). Furthermore, environmental stress could also promote aneuploidy in molluscs (e.g., *M. edulis*, Dixon 1982). In groups of *H. diversicolor* not subjected to cold shock treatments (a triploidy inducer), only 41.8% of the embryos were diploid, 9% were haploids, 7.5% were triploids, and 41.7% were aneuploids (Yang et al. 1998b). Although *H. diversicolor* embryos were only examined at the gastrulae stage (Yang et al. 1998b), triploids can live past the settlement stage. In *H. rubra*, for example, triploids were still viable 22 days after settlement (Liu et al. 2004).

Production of aberrant chromosomes combined with high fecundity and the large variance in marine broadcasts spawners (Hedgecock 1994; Hedrick 2005) could lead to a large number of duplicate loci. The potential presence of duplicate loci in abalone warrants further checking of the patterns of allelic inheritance using known crosses (Selkoe and Toonen 2006). Loci that are present more than twice in a genome can still be included in population genetic studies; however, different analyses would be needed to accommodate the increased number of alleles and the alternative patterns of inheritance (e.g., Bruvo et al. 2004; Kosman and Leonard 2005; Obbard et al. 2006; Luttikhuizen et al. 2007; Kloda et al. 2008)

The other problems that plagued the microsatellite loci developed for this study and prevented their widespread usage included large allele ranges, large allele sizes, and different shaped alleles (Figure 3.3B, C). Potentially, alleles differing by a large size might indicate two different loci were amplified. However, this does not seem to be the case for *H. iris* microsatellites: alleles were at both extreme values and intermediate sizes, and no more than two alleles were present in a sample. Also, large allele size ranges have been recorded for other isolated abalone microsatellites. For instance, locus Hka85 spanned over 200 bp in *H. kamatchatkana* (Miller et al. 2001; Withler et al. 2003), and *H. kamatchatkana* (Miller et al. 2001; Withler et al. 2003), *H. discus hannai* (Li et al. 2002a; An and Han 2006); *Haliotis rubra* (Conod et al. 2002; Baranski et al. 2006) have several loci with allele ranges greater than 100 bp (Appendix 5).

Another problem of screening loci with large size differences between alleles is large allele dropout—the lack of amplification of large alleles due to the preferential amplification of shorter alleles (DeWoody et al. 2006). The larger the difference between alleles the more exacerbated the problem becomes (Björklund 2005). As a result, the large alleles will be missed unless screened with lots of product such that the smaller allele is off-scale. Preferential amplification will lead to over estimates of shorter allele frequencies and

decreased observed heterozygosity (DeWoody et al. 2006). Very large alleles are also difficult to work with because their scoring is limited by the accuracy of the size standard. The largest GeneScan™-500 LIZ® sizing fragment is 500 bp. At this range and beyond, peaks become broader and the lack of sizing fragments make them difficult to score. Potentially, these loci can be screened with a larger size standard, but whether this would efficient in terms cost and time remains to be determined.

Very different shaped alleles may represent amplification of two loci. In the case of Figure 3.2C, the smaller allele had a dinucleotide stutter pattern, and the larger allele had a tetranucleotide stutter pattern. No dinucleotide repeat was found in the AB1 locus sequence (Appendix 6) suggesting that the dinucleotide repeat was a mutant allele or another locus. Furthermore, the shape of the smaller allele in Figure 3.3C is exemplary of how difficult scoring of microsatellites can become (DeWoody et al. 2006; Selkoe and Toonen 2006). The stutter pattern of the smaller allele made deciphering the number of alleles difficult. The dinucleotide stutter of the smaller allele could be considered two alleles; however, this would make a total of three alleles in one individual, which is hard to reconcile with the presumed diploidy of *H. iris*.

After excluding eight loci, the remaining loci were screened across 477 abalone samples. Each locus did not amplify in every individual suggesting either the DNA was poor quality or the presence of null alleles. Null alleles are alleles that fail to amplify because of changes in the primer binding site (DeWoody et al. 2006). The presence of null alleles will lead to a deficit of heterozygotes in the sample data. All three loci were out of Hardy-Weinberg equilibrium, had significant deficits of heterozygotes, and frequencies of null alleles greater than 0.0153. In fact, the frequency of null alleles reported for locus AB14 (0.0644) was not much lower than the high levels reported for oysters (0.094, Hedgecock et al. 2004).

The presence of null alleles could be due to high levels of sequence polymorphisms that may be present in *H. iris* (e.g., see Chapter 6) or abalone in general. Null alleles are not uncommon in abalone: based on segregation analysis of a mapping family, Baranski et al. (2006) had to reject 34 of their 125 microsatellites due to the presence of null alleles. Alternatively, the detection of null alleles may have been biased by the lumping of all samples into a single group. The tests of Hardy-Weinberg equilibrium and estimates of null allele frequencies assume random mating. The lumping of all abalone samples into a single group could have biased these results because *H. iris* are not panmictic and have weak population genetic structure (consistent with Chapter 2). Heterozygote deficits and null alleles should be

reconsidered within non-structured groups (see Chapter 4). Although, the slight genetic structure did not seem to affect tests of linkage disequilibrium, and probably had no affect on tests to detect null alleles.

The loci were very polymorphic with 23 alleles identified for AB14, 84 alleles identified for AB21, and 42 alleles identified for AB31. These numbers were high in comparison to other abalone species but not necessarily unusual. The dinucleotide RubGT1 locus had 41 alleles in 100 individuals of *H. rubra* (Huang and Hanna 1998). Gruenthal et al. (2007) reported 75 alleles in 445 individuals of *H. rufescens* for a locus (Hka3) originally isolated in, *H. kamtschatkana*. In the case of *H. iris*, many of the 84 alleles identified for AB21 varied by one base pair. Further examination of the sequence of locus AB21 (Appendix 6) revealed an eight base pair long poly-A tract, which could be contributing to the high number of single base pair allelic changes. Sequencing of the AB21 locus is needed to confirm this. Alternatively, the abundance of one base pair mutations may result from variation in surrounding temperature of the AB3100 (Davison and Chiba 2003). However, this would be expected to affect all loci. The 42 alleles identified for AB31 were probably real and not due to mis-scoring because many alleles were at the larger end of the size range and consistently differed by four repeats. As a consequence of having so many alleles, the observed heterozygosities were high but not atypical of abalone species (Appendix 5).

The large number of alleles seen in *H. iris* microsatellites makes them more desirable than allozymes (Table 1.1). Precision and accuracy in population genetic studies depend on the sample size, number of loci, level of polymorphism, and the amount of divergence between populations. Despite only three loci, the high levels of polymorphism would aid the precision of estimating genetic distance. Kalinowski (2002) showed that a large number of alleles can be just as useful as a large number of loci when population differentiation is small. In fact when population differentiation is small, increasing sample sizes and/or using loci with higher mutation rates (larger amounts of polymorphism) will increase the precision of estimating genetic distance (Kalinowski 2002, 2005).

Although three polymorphic, relatively scoreable microsatellite loci were identified and screened across 477 individuals, downstream interpretation of this data should be done with caution because the genotyping error rates (per reaction, per allele, and per genotype) were high. Although attempts were made initially to optimize PCRs (including, although not detailed here, the use of additives), lack of time prevented further optimization of PCRs to decrease these error rates. Genotyping errors can result from unreliable PCRs (due to low quantity or quality DNA), null alleles, electrophoresis artifacts, mis-scoring of allele banding

patterns, and human error, i.e., data entry (Hoffman and Amos 2005). The majority of errors seen here were due to mis-entered data.

Due to the large number of errors from poor data entry, all data was rechecked a third time. Assuming this corrected all mis-entered data, then two alleles (one reaction) per locus would be mis-typed giving error rates per allele of 1.2 % for AB14, 1.4 % for AB21, and 1.1 % for AB31. This would also have decreased the error rate per multilocus genotype from 15.3 % to 5.1 %. The impacts of seemingly low error rates per allele (≤ 1 %) are large. Hoffman and Amos (2005) calculate that a "… 1 % error rate in allele calling would lead to almost a quarter of 12-locus genotypes containing at least one error." Assuming the genotyping errors are random, they will introduce noise into downstream population genetic analyses. Such noise will increase the variance surrounding population divergence and mask real patterns that occur.

Ongoing projects

Concurrent to this study, Smith and McVeagh (2006) also employed ATG genetics, Inc. to isolate microsatellites for *H. iris*. ATG genetics, Inc. developed a set of 20 polymorphic microsatellite primers, and from this pool Smith and McVeagh (2006). These markers had between 12 and 48 alleles in 92-93 individuals. The markers deemed unsuitable were subject to the similar problems of amplification of multiple loci and large amounts of allele dropouts and/or null alleles. Along with the unsuitable loci presented here (Table 3.3), these unsuitable microsatellites will be subjected to further optimization attempts. Smith and McVeagh's (2006) suitable primers will be screened across the abalone screened with AB14, AB21, and AB31.

4. Population genetic structure based on three microsatellite loci

ABSTRACT

Although mitochondrial DNA (mtDNA) is useful for studying population genetic patterns, its power to detect structure is limited because it is a single locus that is predominantly maternally inherited. Thus, mtDNA will provide little insight into processes that might be apparent via the incorporation of multiple nuclear genes, such as sex-biased dispersal. In addition, mtDNA analyses might be confounded in some circumstances by patterns of selection that act on the molecule as a whole, and a growing body of literature illustrates that single marker phylogeography based on mtDNA often benefits from the use of additional nuclear markers to confirm or refute the patterns of genetic structure. With few exceptions, population genetic studies of New Zealand marine invertebrates have predominantly used mitochondrial DNA gene sequences as the sole genetic marker. In contrast to the majority of previous studies, this chapter examined the pattern of genetic variation at three microsatellite loci designed for *H. iris,* and compared it to the pattern of genetic structure identified using mitochondrial DNA in Chapter 2. The microsatellite data was similar to the mitochondrial data in that they both identify a north-south split; however, the location of the split was slightly different. Whereas mitochondrial data localized the north-south split to the Cook Strait narrows, the microsatellite data support a split around the regions of upwelling that lie adjacent to Cook Strait.

INTRODUCTION

Mitochondrial DNA is frequently employed in population genetic studies because of its smaller effective population size and faster mutation rates than nuclear DNA (Hartl and Clark 1997 and Brown et al. 1979, respectively). As such, changes in mitochondrial DNA can often be observed within a species, whereas changes in nuclear DNA may not yet be evident (Zink and Barrowclough 2008). However, variation in metazoan mitochondrial DNA does not necessarily reflect genetic variation that occurs or will occur in nuclear DNA. In addition to effective population size and mutation rates, metazoan mitochondrial and nuclear genomes differ in size, structure, location within the cell, copy number, modes of inheritance, replication, segregation, selection pressures, and recombination rates (reviewed in Birky 2001; Ballard and Whitlock 2004; Korpelainen 2004; Xu 2005). As a result, studies report discordance between mitochondrial and nuclear markers in both identifying structure (e.g., Monsen and Blouin 2003; Scribner et al. 2001; Peijnenburg et al. 2006; Borden and Stepien 2006) and estimating F_{ST} (e.g., Lemaire et al. 2005; Caizergues et al. 2003; Sainsbury 1982b).

The lack of recombination in animal mitochondria genomes (although see Kraytsberg et al. 2004) means that the mitochondrial genome is essentially one locus and, regardless of

the number of mitochondrial genes used, genealogies for each gene will be correlated. Increasing evidence suggests that the power of evolutionary analyses lie in the application of multiple unlinked loci (Nei 1987; Edwards and Beerli 2000; Mariette et al. 2002; Rokas and Carroll 2005). Multiple unlinked loci are needed because sampling a single locus or linkage group could misrepresent the distribution of genetic polymorphism within and among genomes. Therefore, studies assessing population genetic structure often employ multiple unlinked nuclear markers, and as technological advances reduce laboratory costs, population genomic approaches are quickly emerging (e.g., Backström et al. 2008)

A number of nuclear markers are available for assessing genetic diversity (e.g., allozymes, randomly amplified fragment length polymorphisms, amplified fragment length polymorphisms, single nucleotide polymorphisms, and intron sequences and are detailed in Baker 2000 and Avise 2004). Marker choice depends primarily on the questions being asked and the amount of time and money being invested (e.g., Gaffney 2000; Sunnucks 2000; Wan et al. 2004). For identifying contemporary patterns of genetic structure, microsatellite markers have become the nuclear marker of choice because they are highly informative (e.g., multiallelic and codominant, Oliveira et al. 2006) and abundant throughout genomes (Tóth et al. 2000).

Molecular markers and the Cook Strait

The identifications of a north-south split in the population genetic structure of New Zealand marine coastal invertebrates have been predominantly based on mitochondrial DNA (Apte and Gardner 2002; Sponer and Roy 2002; Ayers and Waters 2005; Goldstien et al. 2006b; Veale 2007). In a few species, studies were expanded to incorporate nuclear markers. For instance, Sponer and Roy (2002) examined restriction fragment length polymorphisms of the internal transcribed spacer in *Amphipholis squamata*, in addition to mitochondrial 16S rRNA sequences. Since *A. squamata* had a cosmopolitan distribution around New Zealand, they were curious whether the crawling larval stage could sustain gene flow. Both the mitochondrial marker and nuclear marker confirmed distinct lineages of *A. squamata* (putative cryptic species); however, only the mitochondrial marker was used to further examine structure within one of the cryptic lineages.

Nuclear markers have also been employed to study *Perna canaliculus* (Apte and Gardner 2001; Apte et al. 2003; Star et al. 2003). Based on allozymes no north-south split was evident in *P. canaliculus* (Apte and Gardner 2001), but analyses based on mitochondrial DNA supported a north-south split (Apte and Gardner 2002). The discrepancies suggested either

varying processes were differentially affecting the nuclear and mitochondrial genomes or variation in the characteristics of the molecular markers were affecting interpretations of the data. To resolve this pattern, Star et al. (2003) examined the genetic structure of *P. canaliculus* using mitochondrial DNA and randomly amplified polymorphic DNA (RAPD) from the nuclear genome. They found concordant results between the mitochondrial and nuclear markers, which supported the north-south split identified by Apte and Gardner (2002). RAPDs allowed Star and colleagues to sample non-coding genomic regions and more genomic loci and, possibly as a result, to detect genetic structure that was not evident using allozymes.

In addition to allozymes and RAPDs, microsatellites have been applied to examine genetic structure of coastal marine invertebrates around New Zealand. After preliminary findings of no variation in mitochondrial COI and ND5 regions and little variation in the nuclear internal transcribed spacer, arginine kinase intron, and the G-protein coupled receptor, Veale (2007) examined four microsatellites to assess genetic structure of the waratah anemone (*Actinia tenebrosa*). Instead of finding a north-south split, his study detected an isolation by distance pattern. Smith and McVeagh's (2006) preliminary assessment of *H. iris* using six microsatellites found significant differentiation between all sampling sites (two North Island sites, one South Island site, and one Chatham Island site), despite finding no differentiation among North Island and South Island sites using mitochondrial DNA.

The goals of this chapter were to use the data generated in Chapter 3 for microsatellite loci AB14R, AB21R, and AB31R and 1) examine genetic structure in *H. iris* and 2) compare patterns of structure generated from microsatellite loci to patterns of structure generated from mitochondrial DNA in Chapter 2.

MATERIALS AND METHODS

Loci AB14R, AB21R, and AB31R were genotyped across the same 25 samples (477 individuals) used for mitochondrial DNA (Figure 2.3) as described in Chapter 3. Characterization of microsatellites was presented in Table 3.6. Only individuals that could be scored at two or more loci (N = 466) were included in analyses described below.

Population genetic analyses

Variation in *H. iris* loci, AB14, AB21, and AB31, was characterized with standard diversity indices for each sampling locality. The number of alleles (A), the effective number of alleles (Ae), the observed heterozygosity (H_O) and unbiased expected heterozygosity (or

gene diversity, H_E, Nei 1987) were calculated in GeneAlEx 6.1 (Peakall and Smouse 2006), while allelic richness (Ar) was calculated using Fstat 2.9.3 (Goudet 2002). The effective number of alleles accounts for allele frequency as well as allele number to give a more accurate measure of diversity, while allelic richness standardizes the number of alleles by sample size.

Chi square tests are not suitable for testing Hardy-Weinberg equilibrium (HWE) when the sample sizes are small or the expected frequencies are small; instead, randomization tests are preferred (Roff and Bentzen 1989). Deviations from Hardy-Weinberg equilibrium were calculated using an exact test based on a Metropolis-Hastings Markov chain algorithm in GENEPOP 4.0 (Guo and Thompson 1992; Rousset 2008). The default parameters (dememorization length = 10,000, batch length = 100, iterations per batch = 5,000) were used. Gametic disequilibrium between loci within sampling sites was also tested using a Markov chain algorithm in GENEPOP 4.0 (Raymond and Rousset 1995; Rousset 2008). Comparisons between AB21 and either AB14 or AB31 frequently reported "No information" indicating that either the row or column of the contingency table had a marginal sum equal to one (there was no duplicate genotype present for at least one of the loci). An additional test of linkage disequilibrium was carried out using randomizations in Fstat 2.9.3 (1500 permutations at a 5% nominal level, Goudet 2002).

Two different estimates of genetic structure (F_{ST} and R_{ST}) are commonly used when dealing with microsatellites. Even though gene flow might be restricted between demes, the high mutation rate of microsatellites will deflate F_{ST} values (reviewed in Balloux and Lugon-Moulin 2002). In contrast to fixation indices, the R_{ST} estimator (Slatkin 1995) is independent of mutation rate. R_{ST} assumes that microsatellites mutate according to the stepwise mutation model (SMM, Kimura and Ohta 1978), new alleles are created from the addition or deletion of a repeat unit. As a consequence, alleles of similar size are closely related. AB14, AB21, and AB31 sequences contained repeat units of four base pairs, yet alleles differed by one or two base pairs suggesting the SMM was not an appropriate model. Furthermore, the accuracy of R_{ST} is reduced when sampling variance is high (Gaggiotti et al. 1999), mutations deviate from a strict SMM (Slatkin 1995; Balloux et al. 2000), and migration rates are high or little population differentiation exists (Balloux et al. 2000). Given AB14, AB21, and AB31 did not follow a strict SMM and the results from the mitochondrial DNA suggested weak population differentiation, R_{ST} values were not calculated.

F_{ST} as a weighted average over all loci was calculated and tested using 16002 permutations in Arlequin 3.1 (Excoffier et al. 2005). To qualitatively compare the

microsatellite data with the mitochondrial data, a similar set of analyses were performed with the microsatellite data. Relationships among Nei's genetic distance (d_{xy}), calculated in Arlequin 3.1 (Excoffier et al. 2005), were first visualized using metric multidimensional scaling computed in R version 2.6.1 (Team 2007), and the stress was calculated according to Venables and Ripley (1999 p. 333). As a result of a large stress value, further clustering of d_{xy} was performed using the neighbor joining algorithm (Nei and Kumar 2000; Felsenstein 2004) implemented in MEGA4 (Tamura et al. 2007). For mitochondrial DNA, SAMOVAs and haplotype networks were also used to group populations and visualize relationships among haplotypes; however, neither could be used here. SAMOVAs could not be performed on the microsatellite data because the SAMOVA 1.0 (Dupanloup et al. 2002) only calculates R_{ST} values which have already been established as being ill-suited for this data. Since an SMM model is inappropriate, relationships among genotypes cannot be inferred using a network.

A priori groupings (Table 2.2) were tested with Analyses of Molecular Variances (AMOVAs, Excoffier et al. 1992) using 16002 permutations in Arlequin 3.1 (Excoffier et al. 2005). AMOVAs partitioned the total variance of genotypic frequencies into covariance components: among sampling locations (F_{ST}), among sampling locations within groups (F_{SC}), and among groups (F_{CT}). To explore contributions of different samples to genetic structure, pairwise F_{ST} estimates were calculated and tested with 110 permutations Arlequin 3.1 (Excoffier et al. 2005). Additionally, isolation by distance was examined with Mantel tests comparing shortest coastal distances (obtained as described in Chapter 2) and pairwise F_{ST} estimates with 10000 permutations in Arlequin 3.1 (Excoffier et al. 2005). Tests were conducted for all of New Zealand and within a priori groupings.

RESULTS

Sampling locations harbored a lot of variation (Table 4.1, Appendix 7). The number of alleles per sampling locality ranged from five for locus AB14 in sample OCH to 28 for locus AB21 in sample GLN, while expected heterozygosities ranged from 0.716 for locus AB31R in sample NPT to 0.985 for locus AB21 in sample OCH. A large number of alleles were shared between localities: 18 of the 23 AB14 alleles were shared, 66 of the 81 AB21 alleles were shared, and 34 of the 42 AB31 alleles were shared. The large number of shared alleles prevented visualizing the distribution of alleles around New Zealand. Only 12 sampling localities were in Hardy–Weinberg equilibrium across all loci (Table 4.1). Loci AB14 and AB31 were in linkage disequilibrium in samples DBL and NPT ($p = 0.036$ and 0.001,

respectively), while loci AB14 and AB21 were in linkage disequilibrium in sample MTB (p = 0.000).

Genetic structure

Slight but highly significant genetic structure existed across all samples (F_{ST} = 0.009, p = 0.000). Based on pairwise F_{ST}, samples AHU, NPT, and OCH were the most divergent (Table 4.2). AHU was significantly different from ten South Island samples and one North Island sample (OLB). NPT was significantly different from nine North Island samples, one South Island sample (TCL), and the Chatham Islands sample (OCH). The Chatham Island sample was significantly different from two North Island samples and eight South Island samples. After Bonferroni correction only two comparisons were significant (AHU vs. NPT and TIM vs. OCH). To further examine pattern, two clustering approaches were used. First, MDS loosely clustered the samples according to island, except for the North Island sample OLB, which grouped with the South Island samples, and the South Island sample TCL, which grouped with the North Island samples (Figure 4.1). These groupings were also evident in the neighbor joining tree (Figure 4.2).

AMOVAs were used to examine genetic structure across Cook Strait according to a priori groupings proposed in Table 2.2. The largest F_{CT} (0.01505; p = 0.008) was obtained when all North and South Island samples were a group and the Chatham Islands sample (OCH) was a group (Table 4.3). When the Chatham Islands were excluded from the analysis, the overall F_{ST} was still low and highly significant (0.00812, p = 0.000). Excluding OCH, the highest F_{CT} value (0.00916, p = 0.000) was obtained when the north of the South Island samples (PHD and TCL) were included with the North Island samples. To further examine trends among groups, polymorphism was assessed (Table 4.4) Fewer alleles and smaller heterozygosities tended to be found in the South Island groups.

Tests for isolation by distance were significant when all samples were treated as one population and when the North and South Island samples were treated as one population (Table 4.5). No correlations between genetic and geographic distance existed among sampling localities within the groups proposed in Table 2.2, suggesting that the IBD evident among mainland samples resulted from north-south structure indicated with the AMOVAs.

DISCUSSION

Overall, *H. iris* samples were characterized with high levels of variation within sampling localities and very low levels, albeit significant, differentiation between sampling

localities. The large number of samples combined with very polymorphic loci probably aided this study's ability to detect such weak genetic structuring. In general, the structure found using microsatellites supported the structure found using mitochondrial DNA. First, the largest F_{CT} value was produced when the Chatham Islands sample was compared to the North and South Island samples. All but two alleles present in the Chatham Islands sample were shared with other samples: the Chatham Islands had one private allele for AB21 and one private allele for AB31. The presence of shared alleles, yet weak genetic structure, may result from limited gene flow or homoplasy. As discussed in Chapter 1, limited gene flow may result from dispersal by rafting on drifting macroalgae or human mediated transfers.

On the other hand, the shared alleles could have arisen independently among the different samples. The incidences of homoplasy in microsatellite datasets can be quite large and are affected by mutation rate, effective population size, and size-constraints (Estoup et al. 2002). Homoplasy can reduce the number of alleles, the proportion of heterozygotes, and gene diversity, and, ultimately, conceal population substructuring (Estoup et al. 2002). Mutation models, such as the SMM (Kimura and Ohta 1978) and the K-allele model (Crow and Kimura 1970), incorporate homoplasy; however, these were not used because the microsatellites did not appear to be mutating by a gain or loss of a repeat (SMM) and the K-allele model could not be easily implemented in a short time frame.

Also concordant with the pattern that emerged from the mitochondrial data was the structuring among North and South Island samples when the Chatham Islands sample was excluded from the analysis. Testing the north-south split, as in Chapter 2, indicated significant structure around the Cook Strait region, but, unlike the mitochondrial data, the samples PHD and TCL from the north of the South Island grouped with the North Island. Such a grouping was consistent with upwelling (Figure 2.1) creating barriers to dispersal as described in Veale (2007). The discordance of pattern between the mitochondrial data and the microsatellite data may be due to chance or the different mutation rates between markers. Faster mutation rates in microsatellites (reviewed in Schlotterer 2000; Ellegren 2004; Nikitina and Nazarenko 2004; Oliveira et al. 2006) enable these markers, in general for smaller population sizes, to elucidate fine-scale patterns of genetic structure and/or genetic structure on more contemporary time scales, while structure detected using the more slowly evolving mitochondrial DNA reflects structure that has resulted from processes operating in deeper time.

The population genetic structure among mainland samples evident with the microsatellite data was very slight: MDS depicted the relationships among samples as a fairly continuous spectrum of relatedness, the internal branches on the neighbor joining tree were

short, and the F_{ST} (0.00812) was low. Like the mitochondrial data this was inconsistent with the level and pattern of genetic structure identified in *C. ornata* (Φ_{ST} = 0.829, Goldstien 2005, 2006b) and *S. pelliserpentis* (Φ_{ST} = 0.45, Veale 2007). As discussed in Chapter 2, the effects of potential barriers to gene flow in the Cook Strait are species specific and may depend on larval behavior and life history characteristics (Hedgecock 1986; Bohonak 1999; Goldstien 2005; Veale 2007) and/or demographic differences.

The pattern and level (Φ_{ST} = 0.03815) of genetic structure evident with mitochondrial data for *H. iris* was more consistent with species that more intermediate levels of population genetic structure (e.g., *P. regularis* Φ_{ST} = 0.072, Waters and Roy 2004; *P. canaliculus* Φ_{ST} = 0.162, Apte and Gardner 2002; *C. radians* Φ_{ST} = 0.142, Goldstien 2006b). Although the *H. iris* microsatellite data supported a similar pattern, the amount of genetic structure estimated using microsatellites was much smaller. This in part may be related to the high levels of polymorphism in microsatellite markers as opposed to mtDNA. The magnitudes of genetic structure are not comparable because they depend on marker polymorphism: the more polymorphic the marker the smaller the estimates of population structure (Hedrick 1999; Balloux et al. 2000). Furthermore, F_{ST} is based on IAM, which may not be valid for microsatellites, and, hence, differentiation might be underestimated.

In comparison to Veale's (2007) investigation of genetic structure of New Zealand *Actinia tenebrosa* using four microsatellite markers, *H. iris* had a lower level of population genetic structure. However, the *H. iris* microsatellites were more polymorphic with 23–84 alleles per locus than the *A. tenebrosa* microsatellites with 2–17 alleles per locus. Due to clonal replication, the *A. tenebrosa* data also only included unique genotypes for each sample, which would change the allele frequencies and affect downstream analyses. The pattern presented in an MDS analysis of *A. tenebrosa* was continuous like that observed for *H. iris*. A Mantel test indicated an isolation by distance pattern (r^2 = 0.12, $p < 0.0001$) for *A. tenebrosa* comparable with that identified, here, in *H. iris* (r^2 = 0.191, p = 0.019). The isolation by distance pattern in *A. tenebrosa* was further supported with spatial autocorrelation indices (Moran's and Geary's, Sokal and Wartenberg 1983; such indices were not calculated for *H. iris*). However, Veale 2007 did not directly test a split in the Cook Strait region, which may have contributed to these significant finding.

The a priori structures may not represent the true genetic structure of *H. iris*. Cluster analyses indicated that only the TCL sample from the South Island grouped with the North Island samples, while the OLB sample from the North Island grouped with the South Island. Rerunning an AMOVA with these groupings (F_{CT} = 0.01154, p = 0.000) explained more

variation than the a priori groupings. These groupings are similar to the clustering results for the mitochondrial DNA. TCL may be similar to the North Island samples because of its close proximity to the North Island (in particular the WLG sample, Figure 2.3). Clustering of sample OLB with the South Island would be consistent with the northward flowing Wairarapa Coastal Current; however, the geographically intermediate sample MAT grouping with the North Island does not support dispersal via the Wairarapa Coastal Current. This may be a result of stochastic variation, or, alternatively, the mitochondrial DNA data could be reflecting affinities among past populations when haplotypes were shared along the east coast on New Zealand, and this pattern is now changing such that it cannot be traced in more rapid microsatellite markers.

Clustering techniques indicated no patterns of regional structure akin to the regional structure evident in *A. tenebrosa* (Veale 2007). However, the pattern found here was similar to Smith and McVeagh's (2006) preliminary findings for four samples of *H. iris* taken from geographically isolated sites around New Zealand. Their locations included Great Barrier Island, New Plymouth, Stewart Island, and the Chatham Islands. In this study, the samples that were geographically closest to the Smith and McVeagh (2006) samples were OPT, AHU, STR, and OCH, respectively. These samples were differentiated (F_{ST} = 0.01460, p = 0.009), as would be expected because three samples spanned the Cook Strait (OPT and AHU from the north and STR from the south) and one sample came from the Chatham Islands (OCH). Pairwise F_{ST} indicated that each Smith and McVeagh sample was genetically distinct, whereas only two significant pairwise F_{ST} values occurred among OPT, AHU, STR, and OCH: pairwise F_{ST} values between AHU and STR and between STR and OCH were significant (Table 4.2). This inconsistency between studies could result from chance, using different markers, slightly different locations, and/or different samples.

Inbreeding and abalone

In Chapter 3, all three loci were found to be out of Hardy–Weinberg equilibrium. Detection of significant structure among *H. iris* samples violated the assumption of random mating under Hardy–Weinberg equilibrium, potentially explaining the disequilibrium found for all three loci in Chapter 3. In this Chapter, Hardy–Weinberg equilibrium was tested within samples, and significant deficiencies of heterozygotes still existed. Such a finding supported that the grouping of samples in Chapter 3 was not responsible for deviations from Hardy-Weinberg equilibrium and that the deviations probably resulted from the presence of null alleles or preferential amplification as discussed in Chapter 3. Most of the cases of

heterozygote deficiency in *H. iris* samples were restricted to loci AB14 and AB21 (both of which had a larger estimated frequency of null alleles as calculated in Chapter 3).

Heterozygote deficiencies have been found in other abalone microsatellite studies (e.g., *H. kamtschatkana*, Withler et al. 2003; *H. rubra*, Huang et al. 2000 and Conod et al. 2002). Heterozygote deficiencies can result from null alleles, preferential amplification, or inbreeding. Inbreeding should affect all loci, whereas null alleles and preferential amplification should be locus specific. Here, interpretation of heterozygote deficiencies was difficult because deficiencies were both sample and locus specific. Furthermore, global tests of heterozygote deficiency across the three loci were significant for fourteen samples, yet none of the fourteen samples had heterozygote deficiencies in all three loci.

Only four samples (DSD, JCH, TIM, and WST) of these fourteen had heterozygote deficiencies in two loci. Assuming that these loci are not linked to loci under selection, localized inbreeding or substructure could be responsible for the deficiency of heterozygotes in these samples. Although the exact location of the DSD and WST collections were unknown (because they were collected by commercial fisherman), these samples most likely came from within a fiord and an inlet, respectively, where migration to and from the open coast may be sparse (e.g., *Evechinus chloroticus*, Perrin 2002).

In contrast, JCH and TIM samples came from the open coast. The only incoming current on the west coast of the South Island comes from the Tasman Sea (Figure 1.4) and, as a result, may promote local recruitment in the JCH sample. Local recruitment in this region would be consistent with patterns of local recruitment predicted for the *Jasus edwardsi*, which has a much longer (1–2 year) larval stage (Chiswell and Booth 2008). If *H. iris* larvae travel via the Southland Current (Figure 1.4), then the location of TIM, on the convex coastline of the southeast of the South Island, may limit the exposure to this current and promote local recruitment. Further investigations, including more markers, larger sample sizes, and breeding experiments, are needed to further explore the factors causing the heterozygote deficiencies.

5. Variation in gamete recognition proteins

ABSTRACT

Compatibility between sperm and egg recognition proteins are critical for successful fertilization in broadcast marine spawners. Egg recognition proteins identified for sea urchins, mussels, teguline snails, and oysters evolve rapidly, as a result of positive selection, and are variable within species. The abalone egg recognition protein, lysin, also evolves rapidly via positive selection, but intraspecific variation in lysin is apparently rare. However, studies examining intraspecific variation in abalone protein lysin have to date been limited to small samples sizes (N < 11). Given the amount of variability documented in *Haliotis iris* mitochondrial DNA (Chapter 2) and microsatellites (Chapters 4), a subset of the samples used in these previous chapters were screened to examine variability in lysin sequences. A 783 bp fragment of lysin, spanning exon 4–5, was sequenced in 289 individuals from 17 samples around New Zealand. In comparison to mitochondrial and microsatellite markers very little variation was detected. To qualitatively compare the distribution of lysin sequence variation around New Zealand to that of mitochondrial and microsatellites, lysin sequences were subjected to similar tests of genetic structure. In contrast to mitochondrial and microsatellite data, no genetic structure can be inferred from lysin sequences. The lack of variation and structure across sampling locations is consistent with pre-existing ideas of selective sweep and sexual conflict.

INTRODUCTION

Studies of marine differentiation focus on physical (e.g., Addison and Hart 2004; Goldstien 2005; Reid et al. 2006; Lee and Boulding 2007) and behavioral barriers (e.g., Lourie et al. 2005; Bird et al. 2007; Richards et al. 2007; Crandall et al. 2008). However, under both allopatric and sympatric models of speciation, differentiation may result in reproductive isolation such that individuals may disperse and settle in new areas but may not be able to successfully reproduce with the local population. In effect, populations diverge via local adaptation and assortative mating, which in a broadcast spawner is most probably driven by incompatibility between gametes, rather than more overt mate choice (Palumbi 1994).

For marine invertebrates that are broadcast spawners, the interactions between sperm and egg recognition proteins are critical for fertilization (Vacquier et al. 1990; Zigler et al. 2005). These interactions are dependent on the amino acid sequences and, thus, the underlying nucleotide sequences encoding the gamete recognition proteins. As populations differentiate, neutral sequence will change via genetic drift. However, gamete recognition sequences are under intense selection (Geyer and Palumbi 2003; Hellberg et al. 2000; Riginos and McDonald 2003; Springer and Crespi 2007; Moy et al. 2008) and evolve very rapidly (Swanson and Vacquier 2002). Assuming the pattern identified by neutral markers reflects the

true divergence of the population, gamete recognition proteins could variously show a similar pattern of divergence as neutral markers, show no pattern of divergence, or show a more extreme pattern of divergence. These alternatives likely depend on the amount of gene flow, the effective population size, and the strength of selection acting upon the proteins.

Both sperm and egg recognition genes have been identified and sequenced in sea urchins (Kamei and Glabe 2003) and abalone (Fridberger et al. 1985; Galindo et al. 2002), while only egg recognition genes have been identified and sequenced in mussels (Takagi et al. 1994), teguline snails (Hellberg and Vacquier 1999; Hellberg et al. 2000), and oysters (Moy et al. 2008). Most studies examining these genes address their evolution and their role in speciation (e.g., Biermann 1998; Metz et al. 1998a; Zigler and Lessios 2004; Zigler et al. 2005). Within population studies are relatively scarce, and the research (e.g., Geyer and Palumbi 2003; Riginos et al. 2006; Springer and Crespi 2007) that incorporates more than several individuals for each species concentrates on determining the source of positive selection, e.g. sexual conflict (Frank 2000; Gavrilets and Waxman 2002; Haygood 2004) or reinforcement (Dobzhansky 1940).

Nonetheless, various intraspecific patterns emerge from these interspecific studies. Most notably, intraspecific variation in gamete recognition genes is common. Interspecific comparisons of bindin, the egg recognition protein found in sea urchin sperm, indicate that bindin undergoes diversifying selection. Within species, the coding sequence is polymorphic with nonsynonymous and synonymous substitutions (Metz and Palumbi 1996; Palumbi 1999; Geyer and Palumbi 2003). Likewise, interspecific comparisons of the egg recognition protein in mussels, lysin M7, indicate diversifying selection (Riginos and McDonald 2003). Lysin M7 also has considerable diversity within species (Riginos and McDonald 2003; Riginos et al. 2006; Springer and Crespi 2007). The oyster bindin, which agglutinates unfertilized eggs, has diversified by positive selection (Moy et al. 2008). In comparison to sea urchin bindin and mussel lysin M7, oyster bindin is extremely polymorphic within species (Moy et al. 2008). Much of this polymorphism is thought to be due to a repetitive structure that diversifies in size and sequence by recombination and alternative splicing (Moy et al. 2008).

Intraspecific variation in these gamete recognition genes is important because it provides the variation needed for rapid diversification via assortative mating (Metz and Palumbi 1996). The lack of rare genotypes and selection for a single genotype would constrain diversification (Levitan and Ferrell 2006). Maintenance of variation in egg recognition proteins has been attributed to balancing selection created by sexual conflict. For example in sea urchins, bindin diversity is influenced by sex, density, and genotype

frequencies (Levitan and Ferrell 2006). Rare bindin alleles are maintained because they are favored at high sperm concentrations effectively reducing polyspermy. In contrast, common bindin alleles are favored at low sperm concentrations, when polyspermy is at a minimum (Levitan and Ferrell 2006). Although sexual conflict has been directly studied only in sea urchins (e.g., Levitan and Ferrell 2006), Riginos et al. (2006) and Moy et al. (2008) have invoked sexual conflict to explain maintenance of intraspecific polymorphism in mussels and oysters.

Abalone lysin

Abalone gamete recognition proteins are well characterized. During fertilization, sperm release the protein lysin (Vacquier et al. 1990). Lysin interacts with VERL, a receptor protein expressed on the egg vitelline envelope (VE) and nonezymatically creates a hole in the VE via which sperm pass (Vacquier et al. 1990; Lee and Vacquier 1992; Swanson and Vacquier 1998). VERL (Galindo et al. 2003) and lysin (Lee and Vacquier 1992; Swanson and Vacquier 1998) evolve under positive selection. VERL is a massive glycoprotein (3,722 amino acids in *H. rufescens*, Galindo et al. 2002). *H. rufescens*' VERL contains 22 tandem repeats of 153 amino acids (Galindo et al. 2002). The first two repeats are under positive selection, while the other repeats are homogenized via concerted evolution and evolve neutrally (Galindo et al. 2003).

Lysin is much smaller than VERL. It is a monomer protein with 136 amino acids (Vacquier et al. 1990), but occurs as a dimer (Shaw et al. 1995). The current model of lysin and VERL interaction proposes species-specific recognition between the lysin dimer and VERL repeats 1 and 2, dissociation of the weakly associated lysin dimers, and tight binding of the lysin monomers with the VERL repeats (Kresge et al. 2001; Galindo et al. 2003). Approximately, two lysin monomers bind to each VERL repeat (Swanson and Vacquier 1997; Galindo et al. 2003). The majority of interspecific variation in lysin occurs at the amino terminal end (amino acids 2–15, Vacquier et al. 1990; Lee and Vacquier 1992); although variable regions have also been reported at amino acid regions 30–45, 63–87, and 99–136 (Lee and Vacquier 1995).

Like sea urchin bindin, mussel M7 lysin, and oyster bindin, lysin also undergoes positive selection (Lee and Vacquier 1992; Lee and Vacquier 1995; Lee et al. 1995; Yang et al. 2000; Yang and Swanson 2002). Swanson and Vacquier (1998) suggested that VERL changes are driven by sexual conflict or cryptic female choice, while lysin constantly adapts to ongoing changes in VERL sequences. However unlike sea urchin bindin, mussel lysin M7,

and oyster bindin, studies that include multiple individuals per abalone species indicate that intraspecific polymorphisms in lysin are rare (Lee and Vacquier 1995; Metz et al. 1998b; Swanson and Vacquier 1998; Clark et al. 2007). Lee and Vacquier (1995) noted no polymorphisms in the open reading frame and a single transition in the 3′ untranslated region in seven *H. rufescens* spanning 1000 km of Californian coasts. Similarly, Metz et al. (1998b) found little variation in lysin intron/exon sequences obtained for six individuals of *H. rufescens* sampled across 1200 km of coast. Out of 9330 nucleotides screened, only two nucleotide differences were found. The average percent difference between individuals was nil. In contrast, average percent difference for Gα protein intron and mtCOI were 1.4 % and 0.4 %, respectively. Clark et al.'s (2007) analysis of polymorphism across the entire lysin gene sequence (5597 bp) from *Haliotis tuberculata* found only two synonymous substitutions in 18 copies of lysin. Comparisons of lysin with hemocyanin1 and rpL5 introns showed that lysin had reduced polymorphism (Clark et al. 2007).

Metz et al. (1998b) attributed the lower level of lysin polymorphism to an adaptive sweep due to positive selection on gamete recognition in *H. rufescens*. The lack of polymorphism data and a significant excess of rare polymorphisms based on Tajima's *D* and Fu's *F* in *H. tuberculata* led Clark et al. (2007) to also conclude that a recent selective sweep may have occurred. Interestingly, patterns of a selective sweep were evident in both *H. rufescens* and *H. tuberculata*, despite *H. rufescens* being sympatric with six species of abalone (Metz et al. 1998b) and the *H. tuberculata* sample being allopatric (Clark et al. 2007). Sperm competition, sexual selection, and sexual conflict are processes that can occur in allopatry and sympatry; however, these processes maintain variation in sea urchins, mussels, and oysters, while in abalone they apparently have the opposite effect.

Originally, the aim of this chapter was to examine lysin and VERL polymorphism in *H. iris* in a comprehensive population based framework. This project began simultaneously with attempts to assess genetic structure with mitochondrial DNA and microsatellites. Given the prominent genetic structure of species like *Cellana ornata* and *Sypharochiton pelliserpentis* (Goldstien et al. 2006b; Veale 2007, respectively) and the assumed limited dispersal ability of *H. iris* gametes, larvae, and adults (see Chapter 1), *H. iris* had the potential to be highly differentiated at both neutrally and non-neutrally evolving genes including those involved in gamete recognition, such as lysin and VERL. If *H. iris* was differentiated at lysin and VERL, then the possibility to examine hypotheses regarding their coevolution and their potential roles in reproductive isolation among individuals from disparate populations would exist.

Instead, mitochondrial and microsatellite data indicated that *H. iris* DNA was very variable and that the population had weak genetic structuring. Complications amplifying the 5' end of lysin and any portion of VERL limited this investigation to examining 783 bp of lysin, spanning exon 4–5 was compared across 289 individuals from 17 locations around New Zealand. The aims of this chapter were reduced to determining whether *H. iris* lysin was variable, and whether this variability mimicked the structures identified in Chapters 2 and 4. Further, because *H. iris* is evolutionarily distant from other abalone species (Figure 1.2), this investigation provided another clade for comparison of evolutionary patterns in lysin.

MATERIALS AND METHODS

After attempts to amplify multiple regions of the VERL and lysin, the only consistent amplifications were produced with primers 16-4f and 16-12r (Figure 5.1, Metz et al. 1998b). In *H. iris*, these primers amplified an 826 bp fragment spanning the last lysin intron (Metz et al. 1998b). The predicted coding sequence (based on Lee and Vacquier 1995) matched the *H. iris* coding sequence on GenBank (ACCN: L26273), confirming amplification of lysin. Internal primers (LIF and LIR) were designed for sequencing with Primer3 (Figure5.1, Rozen and Skaletsky 2000).

Due to limited time and funding, the lysin fragment was amplified and sequenced in a subset of the samples examined in Chapters 2 and 4. The subset of samples was selected prior to completion of the mitochondrial and microsatellite data set. Originally, the subset consisted of 18 samples chosen for their geographic spread, their location in regards to previous genetic structures found around New Zealand (discussed in Chapter 2), the overall quality of DNA (to allow for easy and successful amplification), and the importance of their location to New Zealand's fishery. Of these 18, WLG was not examined due to lack of time. The amount of differentiation among the 17 remaining samples based on mtDNA and microsatellites were similar to the amount of differentiation among all 25 samples examined in Chapters 2 and 4 (Table 5.1).

Lysin fragments were amplified, visualized, purified, and sequenced according to the methods used for mitochondrial DNA described in Chapter 2, except with a PCR annealing temperature of 53 °C for 30 s and PCR extension duration of 45 s. Sequences were edited with Sequencher™ 4.2.2 (Gene Codes Corporation). Since the lysin fragment was a nuclear gene, heterozygous base pairs were identified with equal overlapping peaks on the chromatograms (Figure 5.2). Sequence alignment was conducted by hand using Se-Al

v2.0a11 (Rambaut 2002), and all variable sites were confirmed by visual inspection of the chromatograms. Sequences were trimmed to a final length of 783 bp.

Haplotype inference

A drawback of using nuclear DNA is that simple amplification produces genotypic data, and haplotype determination can be costly in terms of time and money (discussed in Zhang and Hewitt 2003). Laboratory approaches to convert genotypes to haplotypes include cloning of PCR products, allele-specific amplification, physical isolation of hemizygous templates, genetic isolation of haplotypes, haplotype separation by SSCP, and DGGE (see Zhang and Hewitt 2003 for more approaches). Alternatively, haplotypes can be inferred from genotypes using statistical approaches (e.g., Clark 1990; Stephens et al. 2001; Niu et al. 2002; Wang and Xu 2003; Scheet and Stephens 2006; Zhang et al. 2006; Andrés et al. 2007). Zhang et al. (2001) and Adkins (2004) did not find much difference between haplotyping programs; although Niu et al. (2002), Stephens and Donnelly (2003), and Niu (2004) have reported differences in the programs, which revolve around issues like the degree of population diversity.

Given limited time, budget, and inherent difficulties in working with *H. iris* DNA, *H. iris* lysin haplotypes were inferred using PHASE 2.1.1 (Stephens et al. 2001; Stephens and Scheet 2005). PHASE 2.1.1 uses a Bayesian approach to reconstructing haplotypes that assumes populations are at or near Hardy–Weinberg equilibrium and uses an approximate coalescent prior. PHASE 2.1.1 was chosen because it is more accurate in reproducing true haplotypes (Stephens and Donnelly 2003). Haplotypes were resolved using the default parameters and the default model for recombination rate variation (Li and Stephens 2003).

Population genetic analyses

A set of analyses similar to those performed on mitochondria data were performed on the lysin data. For details of these analyses refer to Chapter 2. In order to assess sequence variation within samples, standard molecular indices were calculated. The number of polymorphic sites, observed and expected heterozygosities, and nucleotide diversity (π, Nei 1987) were computed for each location using Arlequin 3.1 (Excoffier et al. 2005). Exact tests of Hardy–Weinberg equilibrium based on whole haplotypes were also implemented in Arlequin 3.1 (Markov chain steps = 1,000,000, dememorization steps = 100,000, Excoffier et al. 2005).

Haplotype relationships

To quantitatively assess similarity and differences among haplotypes, percent divergences between haplotype pairs were calculated using maximum likelihood settings in PAUP*4.10b10 (Swofford 1998). Maximum likelihood parameters were established in MODELTEST 3.7 (Posada and Crandall 1998). According to Akaike information criterion (AIC, Posada and Buckley 2004), the most appropriate model of sequence evolution was F81 (Felsenstein 1981). The F81 model assumes unequal base frequencies and equal substitution rates. To qualitatively examine haplotypes, relationships were inferred using statistical parsimony as detailed in Chapter 2 (Templeton et al. 1992, implemented in TCS Clement et al. 2000).

Genetic structure

To determine whether genetic structure existed, Φ_{ST} was calculated in Arlequin 3.1 (Excoffier et al. 2005), and significance tests consisted of 16002 permutations. Since no genetic structure was found, further analyses of pattern were not warranted.

RESULTS

A total of 823 bp of the lysin gene was amplified in 289 individuals (2N = 578), but the fragment was reduced to 783 bp to account from primer sequences. The amplified fragment covered 110 bp of coding sequence at the 5′ end, 665 bp of an intron, and eight base pairs of coding sequence at the 3′ end (Figure 5.1). Sequences were variable, and 29 genotypes were identified (Appendices 8 and 9). Of the 783 bp, 65 sites were polymorphic. Nine sites contained transitions, 13 sites contained transversions, and 45 sites contained indels. The 45 sites with indels were distributed among three deletions of 24, 20, and 1 bp. The 24 and 20 bp deletions were in the same haplotype and found in three individuals from South Island samples (GOB, NPT, and TIM), while the 1 bp deletion was in a different haplotype and found in two individuals from the Cook Strait region (IHM and TCL). Pairwise distance between haplotypes ranged from 0.13–0.64%.

Although sequences were variable, almost all the variability was located within the intron and not within the exon. The coding sequence translated into 36 complete amino acids. The coding sequence at the 5′ end of the fragment contained no polymorphic sites and was identical to the *H. iris* sequence on GenBank (positions 343–489, ACCN: L26273). One transversion (at position 486) was located in the coding sequence at the 3′ end of the fragment. This was a replacement substitution that changed the amino acid from an arginine

to a serine. This allele was only found in one heterozygous individual. The remaining variation was all located within the intron portion of the fragment.

PHASE 2.1.1 (Stephens et al. 2001; Stephens and Scheet 2005) resolved 24 different haplotypes from the 29 genotypes. Twelve haplotypes were found in more than one population, while 12 haplotypes were private (Figure 5.3). The two most common haplotypes (haplotypes 1 and 16, Figures 5.4) made up 75% and 18% of all the haplotypes sampled and were present at all locations. As expected from the small amount of pairwise differences, haplotypes only differed by one or two base pair changes, except for haplotype 13, which contained the 24 and 20 bp deletions (Figures 5.4). Expected heterozygosities were variable ranging from 0.1908±0.0928 at GOB to 0.6397±0.0832 at CRW (Table 5.1). Fu's F_s was significant for sample CRW; otherwise, neither Tajima's D nor Fu's F_s were significant for a sample indicating that there was no excess of rare mutations. No genetic structure among the samples was evident (Φ_{ST} = -0.00328, p = 0.654).

DISCUSSION

Little variation was found among *H. iris* lysin sequences. Only one polymorphic site was found in the coding regions; otherwise, coding sequences were identical across sampling locations. The single polymorphism was a clear double peak in the chromatogram; however, it was only sequenced in one direction and awaits confirmation with another primer. The lack of coding sequence variation, in this lysin fragment, suggested that there is no expectation of gamete incompatibility between individuals taken from different locations around New Zealand. However, this analysis only considers coding sequence from lysin's 3' end (exon 4 and exon 5), and more variability may exist in the coding region at the 5' end of exon 1.

Previous studies have shown that the 5' end of lysin is highly variable between species (Vacquier et al. 1990; Lee et al. 1995) and important for the species-specific binding ability of lysin (Lyon and Vacquier 1999). The correspondence between variation at the 3' end (exon 4 and 5) and the 5' end is unknown, and attempts to sequence the 5' end in *H. iris* have been unsuccessful. Clark et al. (2007) was able to sequence most of lysin in nine individuals for *H. t. coccinea* (ACCN: EF660417–EF660434). In *H. t. coccinea*, variation in the 5' end seemed to correspond to variation in the 3' end. The *H. t. coccinea* lysin sequences had a total of 20 polymorphic sites and three sites with indels. The coding sequence only had two synonymous substitutions: one transition was in exon 3 and one transition was in exon 4.

Lysin exon variation

The lack of variation in the coding regions of *H. iris* lysin was consistent with the pattern expected following of a recent selective sweep on this gene, which has been hypothesized to have occurred for *H. rufescens* and *H. t. coccinea* (Metz et al. 1998b; Clark et al. 2007, respectively). Although, interestingly, Tajima's *D* and Fu's F_s did not indicate an excess of rare polymorphisms thought to follow a selective sweep. During a selective sweep, the lysin allele that is most advantageous is swept to fixation in all individuals. The fixation of a single allele may be related to reinforcement, the selection against hybridization (Dobzhansky 1940; Marshall et al. 2002). Reinforcement may be responsible for the positive selection evident in interspecific comparisons of lysin among sympatric species (Lee et al. 1995; Yang et al. 2000) and the patterns of genetic structure found in sea urchin bindin (Geyer and Palumbi 2003) and mussel lysin M7 (Springer and Crespi 2007; although see Riginos et al. 2006). A lysin allele could be advantageous because it limits the hybridization between diverging abalone.

The selective sweep hypothesis makes the most sense when divergent populations of sister species coexist. For instance, *H. rufescens* is sympatric with six other abalone species along the coast of California. As a result, a lysin allele may be advantageous because it limits hybridization between *H. rufescens* and other species. As for *H. iris*, reinforcement could potentially be a factor in the fixation of a lysin allele. *H. iris* occurs around New Zealand with two other abalone species. Yet, *H. iris* is quite distinct genetically from other species of abalone (Figure 1.2) and possibly differentiated 80–111 mya (Coleman and Vacquier 2002) such that enough changes may have accumulated in lysin and other genes to exclude the possibility of hybridization decreasing the need for reinforcement.

However, reduced variation in lysin sequences have also been found in allopatric populations of abalone. Clark et al. (2007) analyzed lysin from an isolated population of *H. t. coccinea*, around 900 km from any other island population of abalone. Lysin sequences in *H. t. coccinea* had reduced variation compared to ribosomal protein L5 intron and hemocyanin 1 intron and a significant number of rare polymorphisms (indicated with Tajima's *D* and Fu and Li's *F*). Since this population of *H. t. coccinea* existed in allopatry, Clark et al. (2007) suggested reinforcement was probably not important in the fixation of *H. t. coccinea*'s lysin sequence.

Alternatively (Clark et al. 2007) proposed sperm competition, sexual conflict, sexual selection, and pathogen attack as modes of selection that operate regardless of whether a population exists in allopatry or sympatry. For sea urchins, mussels, and oysters (Levitan and

Ferrell 2006; Riginos et al. 2006; Moy et al. 2008, respectively), sperm competition, sexual conflict, and sexual selection have been proposed to maintain variation in egg recognition proteins. Whether these maintain variation or promote directional selection will depend on the diversity in the egg receptor for the protein. If diversity exists in the receptor protein on the egg, then diversity in the egg recognition protein would potentially confer reproductive success.

The abalone egg has a vitelline envelope receptor for lysin (VERL) that is variable within species. VERL is a massive glycoprotein located on the egg's vitelline envelope and contains 22 tandem repeats of 153 amino acids (determined in *H. rufescens*, Galindo et al. 2002). Swanson et al. (2001) examined polymorphisms in VERL's 3' end, which included repetitive and non repetitive regions, from Californian abalone (*H. rufescens* and *H. corrugata*). Eight *H. rufescens* sampled from two populations spanning 1200 km had lower levels of polymorphism (π = 0.2 %) than ten *H. corrugata* (π = 1.5 %) from one population. Unlike lysin, polymorphism in VERL was not statistically indicative of a recent adaptive sweep for either species. For *H. corrugata*, two types of VERL were identified. These types differed in amino replacements and indels. Only one out of 11 *H. corrugata* sequenced was heterozygous for the two types. *Haliotis corrugata*'s lysin was polymorphic (π = 0.2 % exon 1, 0.7 % exon 2) but did not split the population into two groups like VERL.

Varying levels of VERL diversity have been found in other Californian abalone. Gruenthal and Burton (2005) sequenced 774 bp of the 5' end of VERL repeat 4 for interspecific comparisons. Of the 561 bp published on GenBank (ACCN: AY817690–AY817692) for three adult *H. sorenseni*, there was one synonymous transition. For three adult *H. k. assimilis* (ACCN: AY817697–AY817699), there was one nonsynonymous transversion. For two *H. k. kamtschatkana* (ACCN: AF490761 and AY817704), there was a 1 bp indel (which would give a completely different coding sequence), a nonsynonymous transversion, nonsynonymous transition, and a synonymous transition. Unfortunately, the level of variation in lysin sequences for these species remains to be determined.

The significance of the polymorphism in VERL is unknown. Both Swanson et al. (2001) and Gruenthal and Burton (2005) sequenced regions of the consecutive repeats. The consecutive repeats evolve quickly by concerted evolution (Galindo et al. 2003) and change neutrally or under slightly purifying selection (Swanson and Vacquier 1998; Swanson et al. 2001). Instead of relying on within population variation to avoid polyspermy (Levitan and Ferrell 2006), polyspermy could be avoided if VERL mutates and homogenizes quickly. In

effect, rapid changes in VERL could cause the lysin to "chase" the ever-changing receptor protein (Swanson and Vacquier 1998; Gavrilets and Waxman 2002).

Unfortunately, the repetitive array is probably not involved in the specificity of lysin binding. The first two repeats evolve independently from the repeats that undergo concerted evolution (Galindo et al. 2002; Galindo et al. 2003). These first repeats appear to be under positive selection, and variation in these repeats is probably uncoupled from the variation in repeats 3–22 (Galindo et al. 2002; Galindo et al. 2003). The initial binding of lysin to VERL is thermodynamically more difficult and probably involves VERL repeats 1 and 2; subsequent binding of lysin is less difficult and probably involves repeats 3–22 (Kresge et al. 2001). The relaxed binding of repeats 3–22 might allow for more polymorphisms in these repeats than in repeats 1 and 2. Intraspecific variation in repeats 1 and 2 are unknown and attempts to sequence them in *H. iris* were unsuccessful.

Egg receptor diversity and rapid evolution are not the only ways females can avoid sexual conflicts (Gavrilets and Waxman 2002). Low polymorphism in lysin may also be promoted because lysin needs to bind to all 22 repeats. Repeats 1 and 2 undergo positive selection and evolve independently from the rest of the repeats that evolve via concerted evolution. As a result, lysin has to be able to bind to repeats that have different sequences and are subject to different evolutionary processes. To do this, lysin might take a compositional "middle ground" (Galindo et al. 2003). Galindo (2003) applied this idea of "middle ground" to reconcile the divergent genotypes of VERL with lysin monomorphism in *H. corrugata* found in Swanson et al. (2001). However, this idea of finding "middle ground" could simply be a result of adapting to all repeats 1–22. The conflicting demands of repeats 1 and 2 vs. repeats 3–22 could "trap" the males in a state of reduced mating success and limit polyspermy (Galindo et al. 2003).

Alternatively, little pressure may exist for VERL to diversify. Pressure for females to diversify is enhanced by population density (Galindo et al. 2003). Diversity at high sperm concentrations is promoted to reduce polyspermy (Levitan and Ferrell 2006). *H. iris* and other abalone species live in aggregations. Aggregations would suggest sperm concentrations are high and increase the likelihood of sperm finding eggs. At these high sperm concentrations, diversity in egg receptors should be promoted. However, fishing data (Figure 1.3, FAO 2000) suggest that abalone population sizes have declined over the last 30 years and aggregations may not be large enough to make polyspermy avoidance an issue.

Varying population densities found in sea urchins corresponds with bindin diversity (Biermann 1998; Levitan 2004; Levitan and Ferrell 2006). Similarly, contrasting gamete

recognition protein variation in abalone species that have maintained large aggregations to those with a limited number may show a different trend in intraspecific variation. The reduced population sizes should also have smaller effective population sizes, which would lead to faster fixation of alleles. However, the declines are quite recent and may yet to have an effect at the level of population genetics. For instance despite the recent declines in *H. kamtschatkana*, Withler et al. (2003) were not able to detect an effect (i.e., disruption of gene flow) at the population genetic level based on microsatellite markers. Small population size and fast fixation for alleles is inconsistent with the amount of diversity evident in *H. iris* mitochondrial DNA, microsatellites, and even the lysin intron.

Lysin intron variation

In contrast to the exons, the noncoding lysin sequences were variable. Again, whether this variation indicated significant variation in other aspects of the molecule remains to be determined. The persistence of two dominant haplotypes might reflect this, but numerous attempts to sequence the rest of the *H. iris* lysin have failed. The levels of polymorphism in the intron sequences were lower than expected based on microsatellites and mitochondrial DNA.

Lack of variation in the intron may be related to it being linked to a gene thought to be under strong selection. The lysin sequence consists of over 5000 bp (Clark et al. 2007), and there are sites under positive selection throughout the length of lysin gene (Yang and Swanson 2002). The intron sequenced here is linked to the preceding 110 bp, which contained no polymorphisms. Furthermore, Yang and Swanson's (2002) analysis of selection pressures among lysin amino acids identified two of the 36 complete amino acids sequenced here as positively selected sites. Only 12 of the 36 residues were also buried sites, while the remaining residues were exposed sites, and exposed sites have an increased nonsynonymous substitution rate (Yang and Swanson 2002).

The relative lack of variation might also be due to the lysin intron being a nuclear sequence and evolving at a much slower pace (explored further in Chapter 6). More variation in *H. iris* lysin intron was evident than was observed previously for the *H. rufescens* intron (Metz et al. 1998a) or *H. t. coccinea* intron (Clark et al. 2007). This may be due to the larger number of samples screened for *H. iris*, or to *H. iris* possessing more neutral variation (evident in mitochondrial DNA and microsatellites) in general.

Population genetic structure

No significant structure was found among sampling localities using lysin sequences. Markers with low polymorphism have limited ability to detect population structure (Kalinowski 2002a), and the lack of variation in lysin probably limited its ability to detect structure across sampling locations. In contrast, mitochondrial and microsatellites were much more variable and were able to detect significant albeit slight genetic structure. The prevalence of two main haplotypes in all locations around New Zealand, suggested that abalone from different locations were capable of breeding: there were no cryptic divergences among *H. iris*. Lysin sequences from mainland New Zealand were also similar to the Chatham Islands, which was distinct based on mitochondrial and microsatellite sequences.

The lack of concordance between lysin and the mitochondrial and microsatellite data may be due to selection pressures on lysin or simply different characteristics among nuclear DNA sequences, mitochondrial DNA sequences, and microsatellite markers. If the latter, then the lack of concordance suggested that a contemporary barrier to gene flow was weak or nonexistent and the genetic structures proposed in Chapters 2 and 4 may actually reflect another process, i.e. historical, demographic, or mutational.

If the population genetic structure was more prominent, then maybe genetic structure in the lysin would be evident. For example Clark et al. (2007) looked at structure between subspecies of *H. tuberculata*, and, as would be expected, F_{ST} values were large and highly significant. The F_{ST} for lysin (0.96, $p << 0.01$) resembled more the F_{ST} for mtCOI (0.91, $p = 0.0028$) than F_{ST} for nuclear rpL5 intron (0.52, $p = 0.015$) and hemocyanin 1 intron (0.19, $p = 0.042$; Clark et al. 2007). It would be interesting to test coding and noncoding lysin variation in species with stronger structure than *H. iris* but weaker structure than subspecies, like *H. asinina* (Imron et al. 2007) or *H. midae* (Evans et al. 2004b). In such cases, sequencing the 5' end of lysin and VERL and additional breeding experiments may help elucidate the roles of reinforcement, sperm competition, sexual conflict, sexual selection, and pathogen attack in the evolution of lysin and VERL.

6. Variation in an intron of the Gα1 protein and its utility for inferring evolutionary processes in lysin

ABSTRACT

Haliotis iris lysin (Chapter 5) is considerably less polymorphic than the mitochondrial DNA and microsatellites examined in Chapters 2–4. Potentially, the relative lack of polymorphism may be due to a recent selective sweep or may just be typical of nuclear DNA, which in general has slower mutation rates than mitochondrial DNA or microsatellites. To distinguish between a selective sweep and the inherent mutation rate of nuclear DNA, 857 bp of the Gα1 intron were amplified in 227 *H. iris* from 14 sampling locations around New Zealand. The Gα1 intron sequences were very polymorphic: a total of 112 haplotypes were identified, and expected heterozygosities ranged from 0.8808–0.9833. In comparison, the *H. iris* lysin fragments had considerably less diversity, consistent with a recent selective sweep. In addition, *H. iris* population genetic structure was assessed using the Gα1 intron. Pairwise Φ_{ST} suggested that the Chatham Islands sample was differentiated from mainland samples (9 out of 13 comparisons were significant); however, this split was not supported when it was tested with an AMOVA ($\Phi_{CT} = 0.05034$, $p = 0.070$). Among the North and South Island samples, AMOVAs identified significant genetic structure with the TCL sample (north of the South Island) grouping with the North Island samples, consistent with both mitochondrial and microsatellite data, and the east coast of the North Island sample, MAT, grouping with the North Island, consistent with only the microsatellite data.

INTRODUCTION

In Chapter 5, variation in lysin sequences was assessed across samples of *Haliotis iris*. In comparison to mitochondrial DNA and microsatellite data, the amount of variation observable in lysin was modest. However, lysin is a nuclear gene and, as such, it does not possess the characteristics of mitochondrial DNA (e.g., smaller effective population size and increased mutation rate), nor does it consist of small tandemly repeated sequences like the hypermutable microsatellites. The lack of variation in the lysin coding sequences and the relatively low amount of sequence variation in the lysin intron may result from a selective sweep promoting homogenization of lysin sequences (Metz et al. 1998b). Alternatively, the amount of variation in the lysin fragment may be typical of *H. iris* nuclear DNA.

Testing selection

Selection affects the amount of intra- and interspecific genetic variation and is generally examined using tests of neutrality that apply either a frequency spectrum or a

comparative approach (reviewed in Nielsen 2005 and Chamary et al. 2006). Frequency spectrum approaches assess the frequencies of different alleles and can be applied within species (e.g., Tajima's *D*, Tajima 1989; Fu's F_s, Fu 1997). These tests assume the population is a constant size and has no structure. Results indicative of a selective sweep include an increase in the fraction of mutations at low frequency, while results indicative of positive selection include an increase in the fraction of mutations at high frequency (Nielsen 2005).

However, the results of such approaches are not clear-cut. First, the results indicative of a selective sweep are similar to results expected for negative selection in that they both increase the fraction of low frequency alleles. Second, such tests can be affected by demographic parameters (e.g., population expansion and selective sweep are indistinguishable with Tajima's *D*, Simonsen et al. 1995). A benefit of a frequency spectrum test like the MacDonald-Kreitman test (McDonald and Kreitman 1991), which examines synonymous substitution (where the substitution results in no amino acid change) and a nonsynonymous substitution (where the substitution results in an amino acid change), is that it is robust in regards to demography; however, this test requires at least two species and cannot detect selective sweeps (Nielsen 2005).

Comparative approaches rely on cross species comparisons between the number and type of substitution (synonymous or nonsynonymous, Nielsen 2005). Synonymous substitutions are regarded as neutral because they should not change the resulting protein, while nonsynonymous substitutions are regarded as non-neutral because they change the resulting protein potentially affecting the protein's "fitness". To gage selection, the numbers of synonymous and nonsynonymous substitutions are measured between sequences and are expressed in terms of the potential for a synonymous or nonsynonymous substitution to occur at a site. The resulting values d_S (the number of synonymous substitutions per synonymous site) and d_N (the number of nonsynonymous substitutions per nonsynonymous site) are then compared: $d_N/d_S > 1$ indicates positive selection, $d_N/d_S = 1$ indicates no selection, and $d_N/d_S < 1$ indicates negative selection (Graur and Li 2000; Nielsen 2005).

Unfortunately, the lysin sequences described in Chapter 5 were all from solely *H. iris*, and contained variation only in the introns, where measurements of d_S and d_N are not applicable. Wong and Nielsen (2004) presented a way to test for selection in noncoding regions, but again this method required comparison to coding sequences with variation. The verified coding sequence of the lysin fragment in Chapter 5 had a $d_S = 0$, making the equations proposed by Wong and Nielsen (2004) undefined.

To help distinguish between confounding effects, such as demography and selection, multiple loci can be applied (e.g., Galtier et al. 2000). In terms of searching for selection within a species, a similar logic is applicable: sequences that might be under selection within a species can be compared to sequences assumed to be evolving neutrally. If introns are evolving neutrally, then comparing exons and introns within a gene would be a potential solution. In the case of the *H. iris* lysin fragment, the intron was more variable than the coding region, suggesting the coding region was under selection; however, linkage of the intron to a gene under selection (selective sweep) would bias the amount of variation in the introns. To determine whether the sequences under selection and/or closely linked to regions under selection, sequences could be compared to *other* neutral nuclear sequences. For example, Metz et al. (1998b) and Clark et al. (2007) both used nuclear gene sequences to help test the hypothesis that the lysin gene product might be subject to strong selection.

Metz et al. (1998b) investigated interspecific variation in lysin among *H. rufescens*, *H. corrugata*, and *H. fulgens* and reported low levels of lysin intron sequence divergence (2.7–6.9%) compared to lysin coding regions (13.7–22.1%). The numbers of synonymous substitution per synonymous site (d_S) calculated for lysin exons were 2.3–5.9 times higher than lysin intron differentiation between species. Between species sequence divergences in Gα1 intron (6.2–8.1%) were similar to those obtained for lysin introns. However within species, variation among sequences from three different lysin introns (concatenated length 1555 bp) was lower than variation observed among sequences of the Gα1 intron (750 bp) and mitochondrial cytochrome oxidase I (mtCOI, 528 bp). Lysin intron sequences, obtained from six individuals potentially spanning 1200 km, only contained two transitions and a poly-G region, while average percent differences were 1.4% and 0.4% for Gα1 intron and mtCOI, respectively. Collectively these data suggest an unusually high rate of change for the lysin gene between species, and an unusual low rate of substitution within a species, which has been suggested to be the consequence of strong positive selection (Nielsen 2005).

Similarly, Clark et al. (2007) assessed polymorphism in *H. tuberculata coccinea* to examine the effect of selection on lysin. They compared the entire lysin sequence (5597 bp) with 491 bp of a ribosomal protein L5 (rpL5) intron, 357 bp of a hemocyanin 1 intron, and 435 bp of mtCOI across 9–11 individuals. Only 20 polymorphic sites were found in the entire lysin sequence. The level of polymorphism in lysin (0.06%) was significantly less than the level of polymorphism in rpL5 intron, hemocyanin 1 intron, and mtCOI (0.37%, 1.37%, and 0.16%, respectively). Additionally, 16 out of the 20 polymorphisms were singleton mutations, which were reflected by both Tajima's *D* and Fu and Li's *F* showing a significant excess of

rare low frequency polymorphisms in lysin, and, thus, a strong indication that this gene may be under selection (Tajima 1989; Fu and Li 1993). This pattern was not observed for rpL5 intron, hemocyanin 1 intron, or mtCOI, which did not differ from neutral expectations (Clark et al. 2007).

Overall, the results from Metz et al. (1998b) and Clark et al. (2007) indicated that the lack of variation in lysin introns was atypical of abalone nuclear DNA, which they concluded was consistent with a recent selective sweep at lysin. In Chapter 5, the lysin intron was variable; however, the variability seemed low compared to the variability observed in mitochondrial DNA (Chapter 2) and three microsatellites (Chapter 4). Furthermore, the lysin intron appeared to fit neutral expectations tested with frequency spectrum techniques (Tajima's *D* and Fu's F_s). However, the lysin intron was situated next to coding sequence that had no variation, and if the coding sequence was under selection and linked to the intron, then the lysin intron should have reduced variation. Rather than comparing the lysin intron to neutral models, this chapter compares lysin intron variation to an assumed neutrally evolving intron, similar to Metz et al. (1998b) and Clark et al. (2007). This chapter documents variation in a Gα1 intron for comparison with the lysin results from Chapter 5. Additionally, this chapter uses the Gα1 intron as another molecular marker to study patterns of genetic structure in *H. iris* for comparison with previous chapters.

MATERIALS AND METHODS

The Gα1 protein is a signal transducing protein that is activated when larvae come in contact with cues for settlement and metamorphosis (Wodicka and Morse 1991 and references therein). The position of an intron between exon 6 and 7 is conserved across mammals, *Drosphila*, and abalone (*H. rufescens*). This intron was chosen based on the availability of sequence data at the start of this project and successful initial amplification with the primers described in Wodicka and Morse (1991).

Sequences were originally amplified using the S1 and S2 primers from Wodicka and Morse (1991). Based on these sequences internal primers were designed with Primer3 (Rozen and Skaletsky 2000) to ensure consistent amplification and sequencing (Table 6.1). Initially, amplifications were to be performed on the same subset of DNA samples used in Chapter 5 (Figure 6.1). However since considerably more sequencing was required for this region, lack of time and funding prevented screening samples CBL, JCH, and MTB. The amount of differentiation among the 14 samples based on mitochondrial DNA and microsatellites were

similar to the amount of differentiation among all 25 samples examined in Chapters 2 and 4 and among the subset of 17 samples used in Chapter 5 (Table 6.2).

Gα intron fragments were amplified, visualized, purified, and sequenced according to the methods used for mitochondrial DNA described in Chapter 2, except with a PCR annealing step of 60 °C for 30 s and PCR extension duration of 45 s. Sequences were edited with Sequencher™ 4.2.2 (Gene Codes Corporation). Since Gα intron is a nuclear gene, heterozygous base pairs were determined with equal overlapping peaks on the chromatograms (similar to scoring heterozygous bases in lysin, Figure 5.2). Sequence alignment was conducted by hand using Se-Al v2.0a11 (Rambaut 2002), and all variable sites were confirmed by visual inspection of chromatograms. Sequences were trimmed to a final length of 857 bp.

Sequence verification

The S1 and S2 primers amplified 1328 bp, thus excluding the primer sites, 48 bp at the 5′ end and 33 bp at the 3′ end coded for 16 and 11 amino acids, respectively. Sequences were verified as *Haliotis* Gα1 by taking the 5′ and 3′ amino acid sequences and comparing them with the sequences provided in Wodicka and Morse (1991). The 5′ and 3′ amino acids perfectly matched the amino acid sequence of *Haliotis* Gα1 protein. Reciprocal BLAST, using the reciprocal best hits approach, indicated the *H. iris* sequence was the *Haliotis* Gα1 ($E = 2x10^{-6}$, Tatusov et al. 1997), while no significant similarity was found between the *H. iris* sequence and the *Haliotis* Gα2 (Wodicka and Morse 1991).

Haplotype inference

As discussed in Chapter 5, haplotypes were inferred using PHASE 2.1.1 (Stephens et al. 2001; Stephens and Scheet 2005). To increase the accuracy of haplotype inference, as many haplotypes as possible were physically deduced prior to using the program. The Gα1 intron fragment contained several indels. Indels were exploited either using allele specific sequencing (Hare and Palumbi 1999), in which primers were designed to match the insertion or deletion and, hence, only amplify that allele, or using CHAMPURU 1.0 (Flot et al. 2006; Flot 2007), in which the double sequence created by an indel is deciphered through determining the indel length and examining the surrounding single-stranded DNA. All remaining haplotypes were then inferred in PHASE 2.1.1 (Stephens et al. 2001; Stephens and Scheet 2005) using the default parameters and the default model for recombination rate

variation (Li and Stephens 2003). The program was optimized following the manual's instruction: consistency was confirmed by examining haplotype frequency estimates and goodness-of-fit measures across five applications of the algorithm with different seeds for the random number generator. The derived haplotype data was used for further genetic analyses.

Population genetic analyses

In order to compare the Gα1 intron with the mitochondrial, the microsatellite, and the lysin intron data (Chapters 2–5), a similar set of analyses were performed. See Chapter 2 for details. To assess sequence variation, standard molecular indices were calculated. The number of polymorphic sites, observed and expected heterozygosities, and nucleotide diversity (π, Nei 1987) were computed for each location as well as for the groups proposed in Table 2.2 using Arlequin 3.1 (Excoffier et al. 2005). Exact tests of Hardy–Weinberg equilibrium based on whole haplotypes were also implemented in Arlequin 3.1 (Markov chain steps = 1,000,000, dememorization steps = 100,000, Excoffier et al. 2005).

Haplotype relationships

To quantitatively assess similarity and differences among haplotypes, percent divergences between haplotype pairs were calculated using maximum likelihood settings in PAUP*4.10b10 (Swofford 1998). Maximum likelihood parameters were established in MODELTEST 3.7 (Posada and Crandall 1998). According to Akaike information criterion (AIC, Posada and Buckley 2004), the most appropriate model of sequence evolution was TVM+I+G (proportion of invariable sites I = 0.7800, and gamma distribution shape parameter = 1.1454). The TVM model is a transversional model that is similar to the general time reversible model, except, instead of having instantaneous rates of change for four transversion and two transitions, it only recognizes four transversion rates and a single transition rate (Rodrígues et al. 1990). In accordance to Chapters 2 and 5, relationships were inferred using statistical parsimony as detailed in Chapter 2 (Templeton et al. 1992, implemented in TCS Clement et al. 2000).

Genetic structure

Similar to the previous chapters the presence of genetic structure was tested using Φ_{ST}, calculated in Arlequin 3.1 (Excoffier et al. 2005), and significance tests consisted of using 16002 permutations. To explore the pattern of genetic structure among *H. iris* samples, pairwise Φ_{ST} and cluster analyses were employed. Pairwise Φ_{ST} values (based on pairwise difference) were calculated in Arlequin 3.1 (Excoffier et al. 2005) and tested with 16002

permutations. As with the mitochondrial data (Chapter 2), no reason existed to assume a past bottleneck (Hedrick 1999), therefore the average number of nucleotide substitutions between haplotypes (d_{xy}, Nei and Li 1979; Nei 1987) were used in the two clustering techniques, metric multidimensional scaling and the neighbor joining method (described in Chapter 2).

As in Chapter 2 and 4 further tests to decipher patterns of genetic structure included analyses of molecular variances (AMOVAs) and isolation by distance. A priori groupings listed in Table 2.2 were tested using AMOVAs implemented in Arlequin 3.1 (Excoffier et al. 2005), and significance tests consisted of using 16002 permutations. Isolation by distance was examined using a Mantel test implemented in Arlequin 3.1 (Excoffier et al. 2005). Correlations between coastal distances (determined in Chapter 2) and pairwise F_{ST} values (Weir and Hill 2002) for each pair of sampling locations were calculated and tested for significance with 10,000 permutations. Similar to the mitochondrial data (Chapter 2), haplotypes were shared among sampling locations suggesting migration had occurred preventing the use linearized F_{ST} values (Slatkin 1991). Mantel tests were performed for all sampling locations within each group defined in Table 2.2.

Neutrality tests

As in Chapter 2, Tajima's *D* and Fu's F_s were used to assess whether populations are in mutation-drift equilibrium (Tajima 1989; Fu 1997). Both Tajima's *D* and Fu's F_s were calculated in Arlequin 3.1 (Excoffier et al. 2005) and tested with 10000 simulations for each sampling location and groups defined Table 2.2.

RESULTS

Overall, the Gα intron was very polymorphic. 857 bp of the *H. iris* Gα1 intron were amplified in 227 of a possible 261 individuals. Sequences were variable with 129 polymorphic sites and 16 indels (totaling 51 sites) ranging from 1–17 bp in length. Of the 86 substitutions, 50 were transitions and 36 were transversions. Each sampling location and all locations treated as a single group were in Hardy-Weinberg equilibrium. Expected heterozygosities were large and ranged from 0.8808±0.0421 in sample NPT to 0.9833±0.0105 in sample IHM (Table 6.3). Nucleotide diversities were small and ranged from 0.007174±0.003889 in CCB to 0.12350±0.006384 in sample IHM (Table 6.3). A total of 112 haplotypes were inferred (Appendix 10): 75 haplotypes were private, while 37 haplotypes were shared across at least two sampling locations (Figure 6.1 and 6.2). Haplotypes were between 0.12–1.77% divergent. The most common haplotype (haplotype 75, Figure 6.2) was found in 60 individuals and present at all sampling locations (Figure 6.1; Appendix 11).

The statistical parsimony found four networks. The largest network contained 108 haplotypes, the second largest network contained two singleton haplotypes, and the two smallest networks each contained a singleton haplotype. The haplotypes present in the three smaller networks were differed from haplotypes in the largest network due to base pair changes and multiple deletions (ranging from 1–10 bp). The largest network indicated that the relationships among the haplotypes were quite complex (Figure 6.3). Either recombination or homoplasy resulted in many cycles. Noticeably, the network consists of two parts. The parts differed by the presence of a 5 bp indel: 56 haplotypes had a deletion and 52 haplotypes had an insertion. Haplotypes with the deletion were more frequent in the northern samples, while the insertion was more frequent in southern samples and the Chatham Islands sample (Figure 6.4). The change in prevalence of the insertion occurred around WST on the west coast of the South Island and south of GOB on the east coast of the South Island.

There was a small but highly significant amount of differentiation among all samples (Φ_{ST} = 0.02666, p = 0.000). OCH appeared to be the most divergent and was significantly different from nine other sampling locations (Table 6.4). All significant pairwise Φ_{ST} comparisons occurred between northern samples (North Island and TCL) and the remaining south Island samples. No pairwise F_{ST} values were significant after Bonferroni corrections. MDS indicated samples OCH and IHM were distinct, while the rest of the samples tended to clump with possibly a north-south split, in which sample TCL grouped with the North Island samples. However, the stress value (0.55) was high indicating that the data could not be depicted in two dimensions. The neighbor joining tree contained long exterior branches and short interior branches (Figure 6.5). The longest interior branch split samples according to North and South Island with TCL from the north of the South Island grouping with the North Island and OCH grouping with the Chatham Islands.

When the a priori groupings were tested (Table 2.2, Table 6.5), no significant difference was found between the Chatham Islands (OCH) the mainland samples (Φ_{CT} = 0.05034, p = 0.070). Significant among group variation was found when mainland samples were split into northern and southern groups (0.03188, p = 0.000). Comparing north of the South Island and North Island samples to the remaining South Island samples and the Chatham Islands sample produced the largest significant Φ_{CT} (0.04062, p = 0.000). The among group component explained 4.06% of the variation, while the within group component explained 95.75% of the variation. When the North and South Island were compared without the Chatham Islands, the comparison between groups explained 3.42% of the variation, while the within groups component explained 96.33%. Examining variation within each of the

AMOVA groups revealed that the northern groupings tended to be more diverse, with larger numbers of haplotypes and higher expected heterozygosities (Table 6.6).

Tests for isolation by distance were significant for when North and South Island samples were grouped together, but not significant within each of the groups (Table 6.7). Neutrality tests were significant for almost all groupings of the North and South Island (Table 6.6) suggesting an excess of rare haplotypes.

DISCUSSION

Gα1 intron sequences were very variable and consisted of a large number of indels ranging in size from 1–16 bp. These indels were useful for deciphering haplotypes and increasing the accuracy of PHASE 2.1.1 (Stephens et al. 2001; Stephens and Donnelly 2003). Metz et al. (1998a) found both indels and direct repeats in *H. rufescens* Gα1 intron sequences. No direct repeats were found in *H. iris* Gα1 intron sequences. In addition to indels, the sequences contained 129 polymorphic sites, and 112 haplotypes were identified. The percent sequence divergence (0.12–2.01%) was similar to that reported for 750 bp of *H. rufescens* Gα1 intron (average divergence 1.4%, Metz et al. 1998a).

Comparison with lysin

This chapter sought to obtain polymorphism data from a portion of the nuclear genome, which was assumed to be evolving neutrally, in order to interpret the lysin polymorphism data collected in Chapter 5. Practically all the variation in the lysin fragments was located in the intron. The amount of variation in the lysin sequence was less than the mitochondrial and microsatellite data analyzed in Chapters 2–4. The lack of variation in the coding sequence suggested it was under selection consistent with the patterns of intraspecfic lysin variation identified for *H. rufescens* and *H. t. coccinea* (Metz et al. 1998a; Clark et al. 2007, respectively). Since the coding and noncoding sequences were adjacent and spanned only 857 bp, they could be assumed to be in linkage disequilibrium. If the lysin sequence has a selective sweep, then the lysin intron sequenced here should have less variation than neutrally evolving sequences. If the coding sequence was evolving neutrally, then lysin fragment should have an equivalent amount of variation as a neutrally evolving sequence. If the coding sequence was undergoing balancing selection, then the lysin fragment will have more variation than a neutrally evolving sequence.

H. iris lysin sequences were much less diverse than the Gα1 intron sequences: the numbers of Gα1 intron haplotypes were about four times greater than the number of lysin

haplotypes per sampling location, and expected heterozygosities based on Gα1 intron were almost twice as large as the expected heterozygosities of *H. iris* lysin. Although *H. iris* lysin sequences were variable and could be used to study population genetic structure, the amount of variability in the lysin fragment was probably not representative of variation in nuclear DNA. The reduced variability in lysin was consistent with the reduced levels of lysin variation found in Metz et al. (1998a) and Clark et al. (2007). Although the results support a selective sweep in lysin, they do not exclude purifying selection (selection against new mutants) as an alternative. Unfortunately, selective sweeps (resulting from positive selection) and purifying selection leave similar signatures of increasing the fraction of mutations that occur at low frequency (Braverman et al. 1995). To distinguish between selective sweeps and purifying selection interspecific comparisons are needed (Nielsen 2005).

Curiously, lysin fragments fit the neutral expectations in Tajima's *D* and Fu's F_S neutrality tests, while Gα1 intron did not fit neutral expectations. The Gα intron had many rare alleles and was consistent with the microsatellite and mitochondrial data. This concordance across markers suggested the pattern reflects a recent bottleneck followed by population expansion rather than expansion. The failure of the neutrality tests to detect selection in the lysin intron may have resulted from violating assumptions of the test like constant population size or no genetic structure.

Unfortunately, this study only examined one other nuclear gene whereas Clark et al. (2007) examined two, rpL5 intron and hemocyanin 1 intron. Potentially, Gα1 intron was not representative of nuclear variation in *H. iris*, and the variation in the Gα1 intron was larger than at a neutral locus. The Gα1 protein belongs to a protein family, as such more than one locus may have been amplified. Potentially, amplifications of the true Gα1 intron and a duplicated Gα1 intron occurred and, as a result, more polymorphisms were identified than would be in either intron alone. All heterozygous bases had only two roughly equal peaks on the chromatograms, which was consistent with the presence of two alleles in an individual. However, this does not exclude the possibility of more than two alleles in an individual. The amplification of a single locus should be verified via cloning. Alternatively, Gα1 intron is linked to a locus undergoing diversifying selection, which would result in increased intraspecific variability (Nielsen 2005).

Genetic structure

The large level of polymorphism in the Gα1 intron probably aided its ability to detect weak structuring around New Zealand, whereas the lower level of polymorphism in lysin

prevented the fragment from identifying weak genetic structure. The results regarding the Chatham Islands sample and divergence from the mainland samples were confusing. Unlike AMOVAs based on mitochondrial DNA sequence and microsatellite variability, grouping samples according to mainland sites and the Chatham Island site did not produce a significant Φ_{CT} ($p = 0.070$). However, the Chatham Islands sample was the most divergent sample based on pairwise Φ_{ST}. The discrepancies between pairwise Φ_{ST} and AMOVAs may be related to artifacts of the AMOVA test, such as the treatment of the haplotype as an allele or the resampling of the data to test significance (the AMOVA results could be considered marginally significant). More individuals from the Chatham Islands should be sampled and sequenced to resolve whether the Chatham Islands is divergent from the mainland samples based on the Gα1 intron.

If the lack of differentiation between mainland samples and the Chatham Islands is real, then the separation of the Chatham Islands from the mainland samples was a recent split, and such a pattern would be consistent with different mutation rates for the markers (Zink and Barrowclough 2008). Microsatellites and mitochondrial DNA would be mutating faster and picking up more recent splits (discussed in Chapter 4). However, this would contradict the inferences based on the genetic structure of the mainland samples.

For the North and South Island, the samples were split around the Cook Strait region in a pattern similar to microsatellites: the TCL sample from the north of the South Island grouped with the North Island samples. This grouping was further evident in both the MDS and neighbor joining analyses. The cluster analyses also indicated that MAT was grouped with the North Island samples similar to microsatellites. However, neither the PHD nor the OLB and EAI samples were included in this study, thus this pattern could not be confirmed.

The concordance between microsatellite and nuclear markers do not fit the relative evolutionary rates of these markers (Zink and Barrowclough 2008). Similarity with the microsatellite markers would suggest that the nuclear DNA either changed faster than the mitochondrial DNA, differential gene flow between male and females occurred, or the results were confounded by chance. Although mitochondrial DNA is often assumed to diverge faster than nuclear DNA (often due to haploid inheritance, smaller effective population size, and faster mutation rates), many exceptions exist: slow rates of mitochondrial nucleotide substitution have been reported for corals (Hellberg 2006) and plants (Wolfe et al. 1987). Further investigation is warranted to determine whether abalone mitochondria may also be an exception. In addition, the lack of concordance between nuclear and mitochondrial DNA could have resulted from differential gene flow between the sexes; however, this seems

unlikely because *H. iris* are broadcast spawners with an assumed passive pelagic larval stage. Finally, the lack of concordance may be due to chance given that the genetic structure, identified by both nuclear and mitochondrial markers, was slight. Interestingly, the Gα1 intron showed the same trend as the mitochondrial data with a lot more variation in the northern samples than the southern samples. This trend may also have occurred in the microsatellite data, the high level of polymorphism in the microsatellites versus the sampling under taken could mean that the signal available in those data was not fully resolvable.

7. Summary and Conclusions

In order to understand connectivity of marine populations, it is important to identify present and past barriers to gene flow. This thesis set out to examine connectivity of an endemic New Zealand abalone, *Haliotis iris*, by employing molecular markers to identify genetic structure. The framework used to assess genetic structure centered on a potential extrinsic barrier to gene flow, the Cook Strait region, and a potential intrinsic barrier to gene flow, the gamete recognition protein, lysin.

Cook Strait

The Cook Strait region corresponds with genetic splits identified in other coastal marine invertebrate fauna (Apte and Gardner 2002; Star et al. 2003; Waters and Roy 2004; Ayers and Waters 2005; Goldstien et al. 2006b; Veale 2007). The genetic split has been hypothesized to result from restricted gene flow either due to the recent complex hydrography (i.e., upwelling) or the past oceanographic and geologic history of the Cook Strait region. The benefit of using multiple molecular markers rests in the ability to decipher between what is actually occurring in a population versus what is occurring as a result of the characteristics of the molecular marker. Genetic structuring of *H. iris* around the Cook Strait was explored using three different types of molecular markers: two regions of the mitochondrial genome (totaling 1055 bp), three microsatellite loci, and two regions of the nuclear genome (totaling 1640 bp). The mitochondrial DNA, the microsatellite loci, and one nuclear region, the Gα1 intron, all indicated weak but highly significant genetic structure around the Cook Strait region. These markers were characterized by high levels of polymorphism, whereas the other nuclear DNA marker, the lysin fragment, was in comparison less polymorphic, potentially limiting its ability to detect population structure.

In comparison to other New Zealand marine invertebrates, the amount of genetic structure among mainland *H. iris* samples was small, and the pattern of genetic structure was not clear-cut. Φ_{ST}/F_{ST} values were lower than those for species with prominent genetic structure around Cook Strait (e.g., *Cellana ornata*, Goldstien 2005, 2006b; *Sypharochiton pelliserpentis*, Veale 2007) and those for species with intermediate structure around Cook Strait (e.g., *Patriella regularis*, Waters and Roy 2004; *Perna canaliculus*, Apte and Gardner 2002; *Cellana radians*, Goldstien 2005, 2006b). Differences such as these highlight that the influence of the Cook Strait region on population genetic structure was species specific and

probably depended on differences in larval behavior and life history characteristics (Hedgecock 1986; Bohonak 1999; Goldstien 2005; Veale 2007) and/or demography.

In general, all markers supported a genetic split in the Cook Strait region; however, the position of the split varied slightly according to marker. AMOVAs indicated that more variation in mitochondrial DNA was explained when samples were partitioned across Cook Strait narrows (Table 2.2), while more variation in microsatellites and the Gα1 intron were explained when samples were partitioned such that the north of the South Island samples (PHD and TCL for microsatellites and TCL for Gα1 intron) were grouped with the North Island. The partitioning of the northern South Island samples with the North Island samples would be consistent with other New Zealand coastal marine invertebrates and recent hydrography acting as a barrier to gene flow (as summarized in Veale 2007).

On the other hand, cluster analyses indicate that the a priori hypotheses may not be the best explanation of genetic structure (Figure 7.1). For instance, the variation in mitochondrial DNA may be better explained if the TCL sample (from the north of the South Island) grouped with the north island samples, while the PHD sample (from the north of the South Island) and MAT, OLB, and EAI (from the east coast of the North Island) grouped with the South Island. Such a scenario would be consistent both with recent hydrography and the geological history of New Zealand. The close proximity of the TCL sample to the North Island WLG sample combined with large amount of mixing in the Cook Strait region (Harris 1990) could enable dispersal and, therefore, gene flow. Dispersal from the southern abalone to the east coast of the North Island may also be enabled by the northward flowing Wairarapa Coastal Current (WCC in Figure 1.4, Chiswell 2000). Connectivity along the eastern coast the North and South Islands could also be inferred from New Zealand's geological history. Prior to the opening of the Cook Strait, the North and South Islands were a single continuous landmass in the Pleistocene (Figure 7.2, Fleming 1979). Dispersal could then occur by a stepping-stone process along the coast.

AMOVAs based on microsatellites indicated that the genetic split lied to the south of samples from the north of the South Island (PHD and TCL, Figure 7.1). However concordant with the mitochondrial data, cluster analyses identified an alternative structure: samples TCL, MAT, and EAI grouped with the North Island samples, while samples PHD and OLB grouped with the South Island scenarios. Such a scenario was harder to explain by recent hydrography, and potentially more sampling along the northeast coast of the North Island, applying more molecular markers, or field investigations *H. iris* dispersal along the coast may help explain this pattern.

In contrast to both mitochondrial DNA and microsatellites, the cluster analyses using Gα1 intron data matched the AMOVA groupings, possibly because samples OLB, PHD, and EAI were not screened with the Gα1 intron. If the differences between datasets described above are not due to chance, then the concordance across microsatellite markers and a nuclear gene and not mitochondrial DNA was inconsistent with the characteristics of these markers (Zink and Barrowclough 2008). In general, mitochondrial DNA has a faster mutation rate and a lower effective population size than nuclear DNA; therefore, it should detect more recent structure. However, microsatellites mutate extremely fast in comparison to mitochondrial DNA and should detect more recent structure than mitochondrial DNA (reviewed in Schlotterer 2000; Ellegren 2004; Nikitina and Nazarenko 2004; Oliveira et al. 2006). Therefore, the similarity in structure, deduced from Gα1 intron sequences to that deduced from microsatellites, was not expected and suggested either that the nuclear Gα1 intron changes faster than the mitochondrial DNA in order to be comparable to the microsatellite data or that the microsatellites change slower than the mitochondrial DNA in order to comparable to the Gα1 intron. With as many as 84 alleles identified at microsatellite locus AB21, the latter seems unlikely. The evolutionary rate for the Gα1 intron is unknown, but it is conceivable that this gene region does evolve at a rate as high, if not higher, than the mitochondrial gene regions examined (ATPase6–8 and mtCOI).

Finally, molecular diversities, based on mitochondrial DNA and the Gα intron, were higher in the northern samples than the southern samples. Decreased mutation rates or increased fishing rates may be affecting southern samples. Interestingly, the same trend of increasing diversity in the northern samples was also observed in *Patriella regularis* (Ayers and Waters 2005), which has not been subjected to intense fishing pressures like *H. iris*. Potentially, the warmer water of the North Island (Figure 2.11) could result in higher mutation rates for northern samples and, therefore, an accumulation of more diversity (Gillooly et al. 2005). In contrast, southern samples of *Perna canaliculus* harbored more genetic diversity than northern samples (Apte and Gardner 2002). This seemingly contradicts the hypothesis that warmer water is related to increased mutation rates, yet *P. canaliculus* data is confounded by human-mediated transfers of mussel spat from the South Island to the North Island (Apte et al. 2003).

Lysin

Variation among *H. iris* samples was also examined using lysin. Lysin is the egg recognition protein that is critical for successful fertilization of abalone eggs. Comparison

between species has indicated that lysin is under intense positive selection (Lee and Vacquier 1992; Lee and Vacquier 1995; Lee et al. 1995; Yang et al. 2000; Yang and Swanson 2002). Potentially, migration into a population with different lysin sequences will limit the effects gene flow because the migrant will not be able to reproduce in the new population. Egg recognition proteins tend to be variable within species; however based on the rather limited intraspecific studies of lysin, all of which had sample sizes of 11 or fewer individuals, lysin appears to have little intraspecific variation (Lee and Vacquier 1995; Metz et al. 1998b; Swanson et al. 2001; Clark et al. 2007). Selective sweeps are potentially driving the homogenization of lysin. Given the presence of slight genetic structure, a lysin fragment was examined in 287 individuals from 17 samples around New Zealand to determine if lysin might also be diverging.

Practically no variation was found in the lysin coding sequence; while modest variation was found in the intron sequence. The overall variation found in lysin was reduced compared to the Gα intron. The pattern of reduced variation in a linked region and no variation in the coding region was consistent with hypotheses of selective sweeps (Metz et al. 1998a; Clark et al. 2007). If the lack of variation in the lysin fragment was representative of the remaining lysin sequence, then it seems likely that differentiation in the lysin gene product, altering gamete recognition, was not preventing gene flow between abalone from different localities around New Zealand.

Genetic variation in H. iris

Notably in this study the amount of intraspecific variation in *H. iris* was large. The variation in *H. iris* appeared to be larger than that detected in other New Zealand coastal marine invertebrates. For instance, expected heterozygosity for *P. canaliculus* around New Zealand ranged from 0.0844–0.3543 for seven allozymes (Apte and Gardner 2001), and haplotype diversity ranged from 0.6572±0.1329 for mitochondrial NADHIV (Apte and Gardner 2002). Haplotype diversity of cytochrome b for *C. ornata* and *C. flava* samples from around New Zealand were 0.638±0.017 and 0.333 ±0.056, respectively (Goldstien et al. 2006b).

Potentially, the lower levels of variation in these species were related to the molecular marker used. Allozymes are notorious for not being very polymorphic and potentially under selection. Although the mitochondrial genome is essentially one locus, regions of the mitochondria do not necessarily change at the same rate. With the *H. iris* data set, the mtCOI region was less polymorphic and had less haplotype diversity than the ATP8 region. Ayers

and Waters (2005) identified 132 haplotypes in 284 individual of *P. regularis*. This large number of haplotypes might be due to the inclusion of the mitochondrial control region, a hypervariable region of the mitochondria. Alternatively, the high levels of variation detected within *H. iris* may be related to effective population size. As opposed to *Cellana* spp., *H. iris* occurs within the subtidal and potentially has access to more habitat than *Cellana* spp., which are restricted to the intertidal.

In comparison to other abalone species, large levels of intraspecific variation are not uncommon. For *H. cracherodii* (Gruenthal and Burton 2008), *H. rufescens* (Burton and Tegner 2000), *H. midae* (Evans et al. 2004b), and *H. rubra* (Conod et al. 2002), mitochondrial haplotype diversities ranged from 0.630–0.8918 (Table 1.2). *H. asinina* had the largest range of haplotype diversities—from a sample of seven with identical mtCOII haplotypes to a sample of nine with eight different mtCOII haplotypes (h = 0.9722). Isolated microsatellites for abalone also have large expected heterozygosities. For instance, *H. kamstschatkana* expected heterozygosities ranged from 0.91–0.93 per sampling location (Withler et al. 2003). Although, the range of heterozygosities for microsatellites were a lot more variable spanning from 0.414 in *H. rubra* (Temby et al. 2007) to 0.867 in *H. fulgens* (Gutiérrez-Gonzalez and Perez-Enriquez 2005).

Intriguingly, one of the most extreme cases of genomic variation is the urochordate *Ciona savignyi*, an ocean-dwelling broadcast spawning sea squirt (Small et al. 2007). The large amount of variation in *C. savignyi* was deduced to be due to large effective population size and not due to an increase in mutation rate. Abalone could equally have large effective population sizes, although the intense fishery practices suggest variability should be somewhat reduced. Lack of evidence for reduced variability due to fishing (e.g., Withler et al. 2003) may simply be due to a lag time between a demographic event and when that demographic event is detected by molecular markers.

Potential problems and future investigations

The study of *H. iris* population genetic structure had weaknesses that could be corrected or considered before future studies. First due to funding, time, a lack of resources, and post hoc identification of surprisingly high genetic diversity in this species, a less than ideal number of individuals were taken from each location. This limited the precision of genetic distance measures. Accounting for the structure identified here and the structures identified in other studies of New Zealand coastal marine invertebrates, future studies should aim to either collect more samples per location, pool samples, use more variable markers,

and/or use more molecular markers. Second although this study identified alternative structures not proposed a priori, the accuracy of these structures was debatable: all MDS analyses had high stress values and the neighbor joining algorithm can be spurious. Future studies could sample to test these alternative structures as a priori structures. Third, the microsatellite results should also be interpreted cautiously because the error rate per genotype was high (15.3 % per multilocus genotype; Hoffman and Amos 2005). Future studies could apply a different set of microsatellites or work to further optimize the microsatellites used here. Application of more microsatellites would also enable different analytical approaches to be considered, such as Bayesian clustering analyses (Pritchard et al. 2000) and landscape genetic approaches (Miller 2005; Chen et al. 2007).

The study of *H. iris* lysin variation also presented several issues that could be examined or considered in future investigations. The region of lysin sequenced here was towards the 3′ end of the protein and may not be representative of variation at the 5′ end of the protein (which is involved in species specific interactions). Future studies of lysin may want to examine different regions of lysin to see if they differ in diversity. The diversity in the lysin sequences was also compared to the diversity in the Gα1 intron to determine if lysin sequence was less variable than a neutral sequence. Future studies may want to apply other neutral sequences to confirm that the Gα1 intron is evolving according to neutral predictions. They also should consider sampling different regions of the exons of lysin in other species to apply different tests of selection.

Despite the above weaknesses, the data accumulated in this study are beneficial to future abalone studies. First, this data provides a baseline for comparative analyses with the other New Zealand abalone, *H. virginea* and *H. australis*. Comparison with these other species would be interesting because they differ in their locations on the rocky reefs (e.g., *H. virginea* occurs in crevices and under boulders), their behaviors (e.g., *H. australis* does not occur in aggregations), and their distribution to offshore islands (e.g., *H. virginea* and *H. australis* occur on New Zealand subantarctic islands as well as the Chatham Islands). Given New Zealand's diverse, yet isolated, marine environment and the potentially large *H. iris* population size, the data presented here could also act as baseline data for time-series or comparative studies to determine the role of selection in shaping population genetic structure (Lenormand 2002; Kawecki and Ebert 2004; Nielsen 2005).

This study has isolated markers for further development and application in the abalone aquaculture industry, i.e. genetic and heritability studies. Countries like Australia and Japan have already developed large microsatellite data sets for genetic mapping projects (e.g.,

Baranski et al. 2006; Sekino et al. 2006, respectively), and commercial interests in New Zealand have already started to incorporate the microsatellites described here in their current research. In addition to industry, this research may also benefit fisheries management. The driving force of many abalone population genetic studies is stock assessment. Now that the data is being collected, a critical assessment is warranted on how fisheries should incorporate this data into future management decisions and modeling stock viability. Given the data accumulated here, the plethora of research on *H. iris* biology, and over 20 years of documented management, *H. iris* seems like an ideal model for assessing the application of genetic data to fisheries management.

Tables and Figures

Figure 1.1: Distribution of *Haliotis* spp. 114
Figure 1.2: Abalone phylogenies 117
Table 1.1: Summary of abalone population genetics research that used nuclear data 118
Table 1.2: Summary of abalone population genetics research that used mitochondrial data 120
Figure 1.3: World abalone catch 121
Figure 1.4: Ocean currents and rocky reefs around New Zealand 123
Table 2.1: Summary of New Zealand marine invertebrate population genetic research 124
Figure 2.1: Cook Strait region 126
Figure 2.2: Amplified fragments of *H. iris* mitochondrial DNA 127
Figure 2.3: Locations and sample sizes for *H. iris* collected around New Zealand 128
Table 2.2: A priori population genetic structures 129
Table 2.3: Polymorphism data and neutrality results for mtCOI, ATPase8–ATPase6 130
Table 2.4: Polymorphism data and neutrality results for concatenated mitochondrial fragments across sampling locations 131
Figure 2.4: Mitochondrial haplotype frequencies and sample sizes at collection locations around New Zealand 132
Figure 2.5: Minimum spanning network based on mitochondrial DNA 133
Figure 2.6: Statistical parsimony network based on mitochondrial DNA 134
Figure 2.7: Median joining network based on mitochondrial DNA 135
Figure 2.8: Multidimensional scaling based on mitochondrial DNA 136
Figure 2.9: Neighbor joining analysis based on mitochondrial DNA 137
Table 2.5: Spatial analysis of molecular variance (SAMOVA) based on mitochondrial DNA 138
Table 2.6: AMOVA results based on mitochondrial DNA 139
Table 2.7: Polymorphism data and neutrality results for mitochondrial DNA across AMOVA groupings 140
Table 2.8: Pairwise Φ_{ST} based mitochondrial DNA 141
Figure 2.10: Nested network for Nested Clade Phylogeographic Analysis (NCPA) 142
Table 2.9: Interpretation of Nested Clade Phylogeographic Analysis 143
Table 2.10: Mantel tests using mitochondrial DNA 143
Figure 2.11: Mean sea surface temperatures around New Zealand from 1993–2002 144
Figure 2.12: *Haliotis iris* fishing management areas and Total Allowable Commercial Catch (TACC) per area for the 2008–2009 fishing season 145
Table 3.1: Loci used for testing cross-species amplification on *H. iris* DNA 146
Figure 3.1: Example electropherogram for a microsatellite labeled with dUTPs 147
Table 3.2: Primer pairs for polymorphic microsatellite loci isolated by ATG genetics, Inc. 148
Table 3.3: Preliminary screening of 13 *H. iris* microsatellite loci 149
Table 3.4: Optimized PCR conditions for loci AB14, AB21, and AB31 149
Figure 3.2: Examples of electropherograms for loci AB14, AB21, and AB31 150
Figure 3.3: Examples of electropherograms for loci deemed unsuitable for further screening 151
Table 3.5: Genotyping error rates 152
Table 3.6: Characterization of loci AB14, AB21, and AB31 152
Table 4.1: Standard polymorphism indices for 25 sampling locations and three microsatellite loci 153
Table 4.2: Pairwise F_{ST} based on microsatellites 154
Figure 4.1: Multidimensional scaling based on microsatellites 155
Figure 4.2: Neighbor joining analysis based on microsatellites 156

Table 4.3: AMOVA results based on microsatellites 157
Table 4.4: Standard polymorphism indices for AMOVA groupings based on microsatellites.. 158
Table 4.5: Mantel tests using microsatellites 158
Figure 5.1: Overview of lysin 159
Figure 5.2: Examples of heterozygous base calls for nuclear sequences 160
Table 5.1: Comparison of the amount of differentiation between the samples used in Chapters 2 and 4 and those used in Chapter 5 160
Figure 5.3: Lysin fragment haplotype frequencies and sample sizes at collection locations around New Zealand 161
Figure 5.4: Statistical parsimony networks based on lysin fragments 162
Table 5.2: Polymorphism data and neutrality results for 783 bp of lysin across 17 locations around New Zealand 163
Table 6.1: Gα1 intron primers 164
Table 6.2: Comparison of the amount of differentiation between the samples used in Chapters 2, 4, and 5 and those used in Chapter 6 164
Table 6.3: Polymorphism data and neutrality results for 857 bp of Gα1 intron across 14 locations around New Zealand 165
Figure 6.1: Gα1 intron haplotype frequencies and sample sizes at 14 locations around New Zealand 166
Figure 6.2: Statistical parsimony networks based on Gα1 introns 167
Figure 6.3: Frequency of a 5 bp deletion at 14 locations around New Zealand 168
Table 6.4: Pairwise Φ_{ST} based on Gα1 intron 169
Figure 6.4: Multidimensional scaling based on Gα1 intron 170
Figure 6.5: Neighbor joining analysis based on Gα1 intron 171
Table 6.5: AMOVA results based on Gα1 introns 172
Table 6.6: Polymorphism data and neutrality results for Gα1 intron sequences across AMOVA groupings 173
Table 6.7: Mantel tests using Gα1 introns 174
Figure 7.1: Comparison of AMOVA and cluster analyses 175
Figure 7.2: Reconstruction of the New Zealand land mass over the last 12 Myr 176

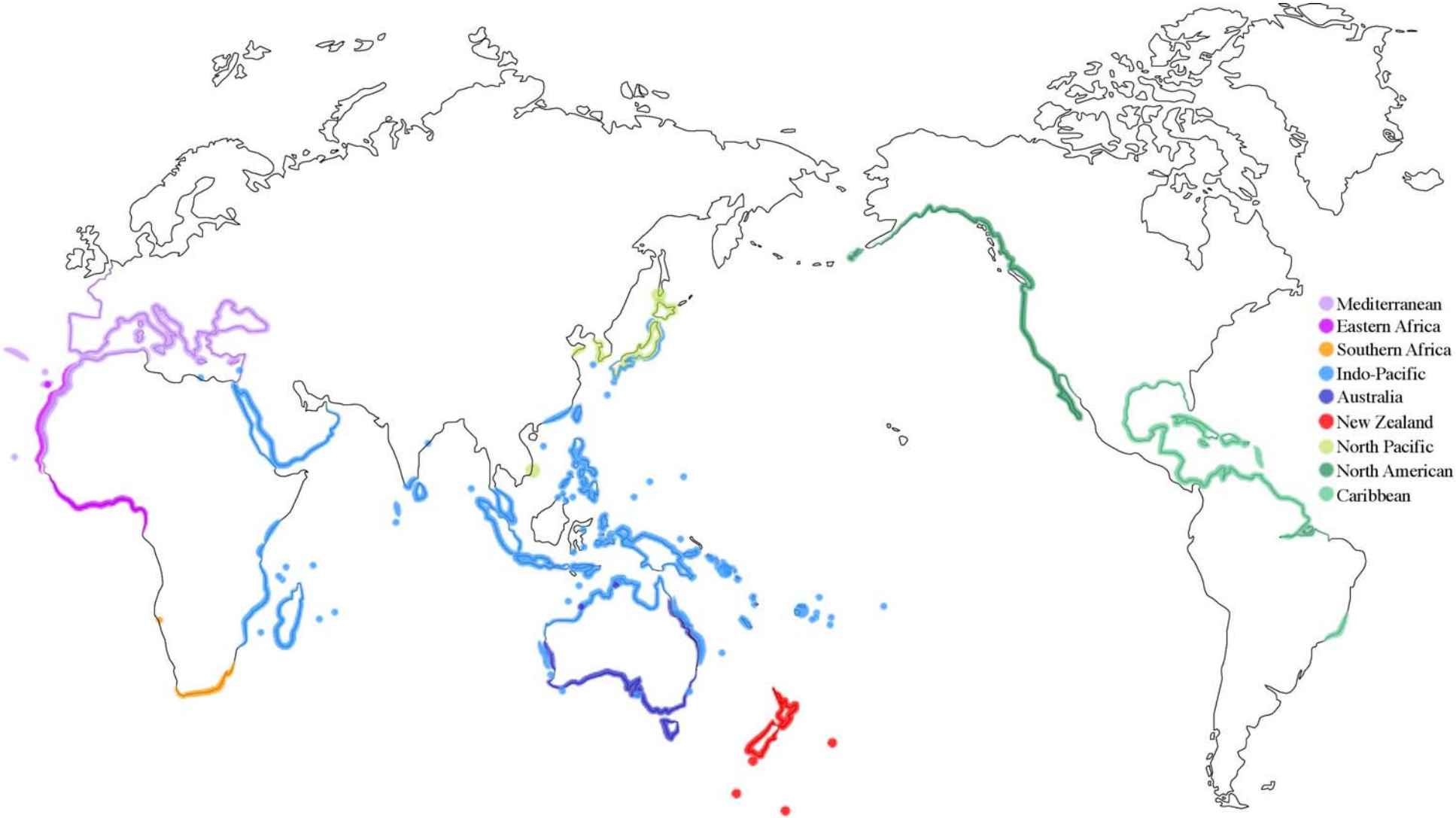
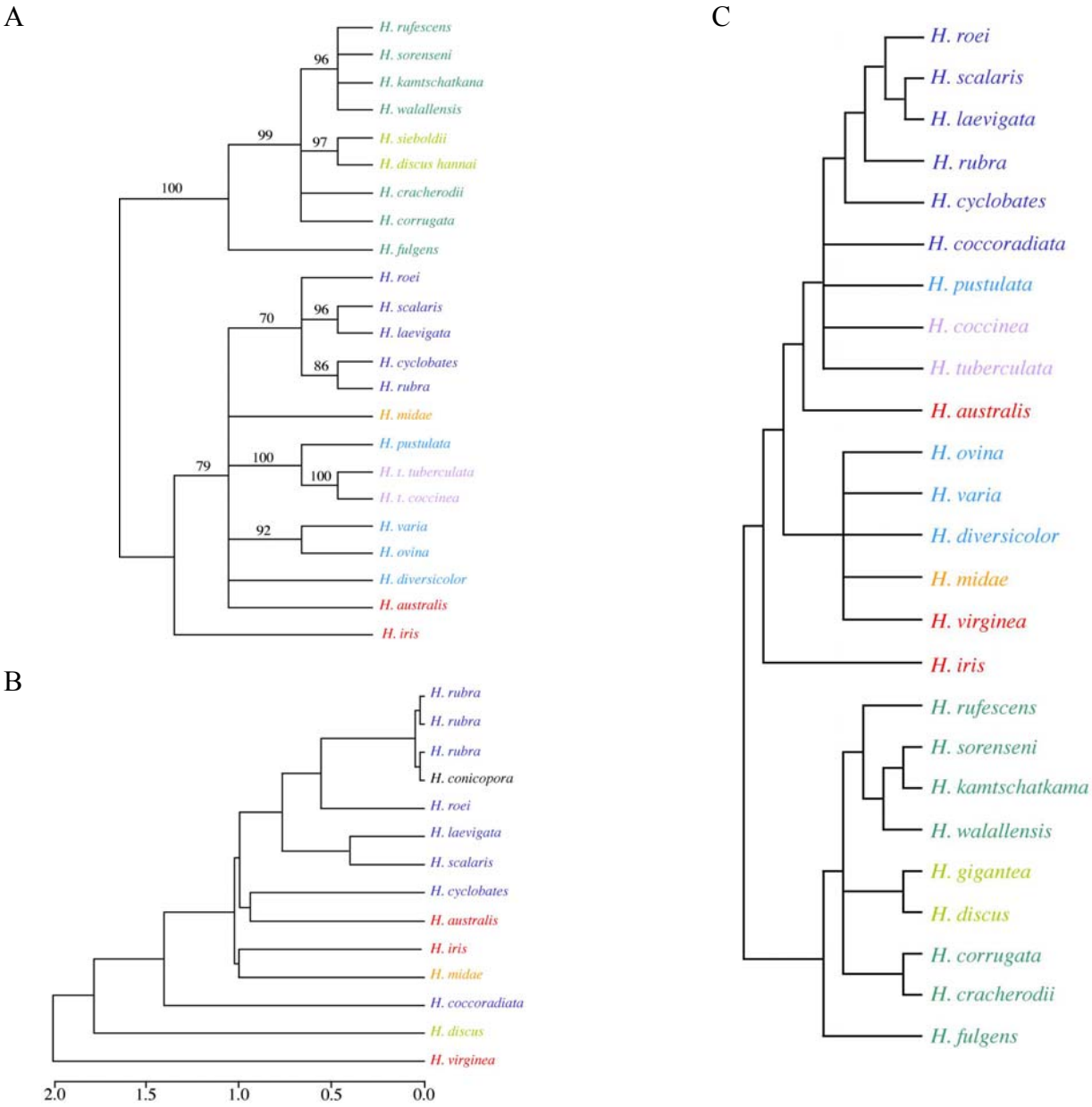
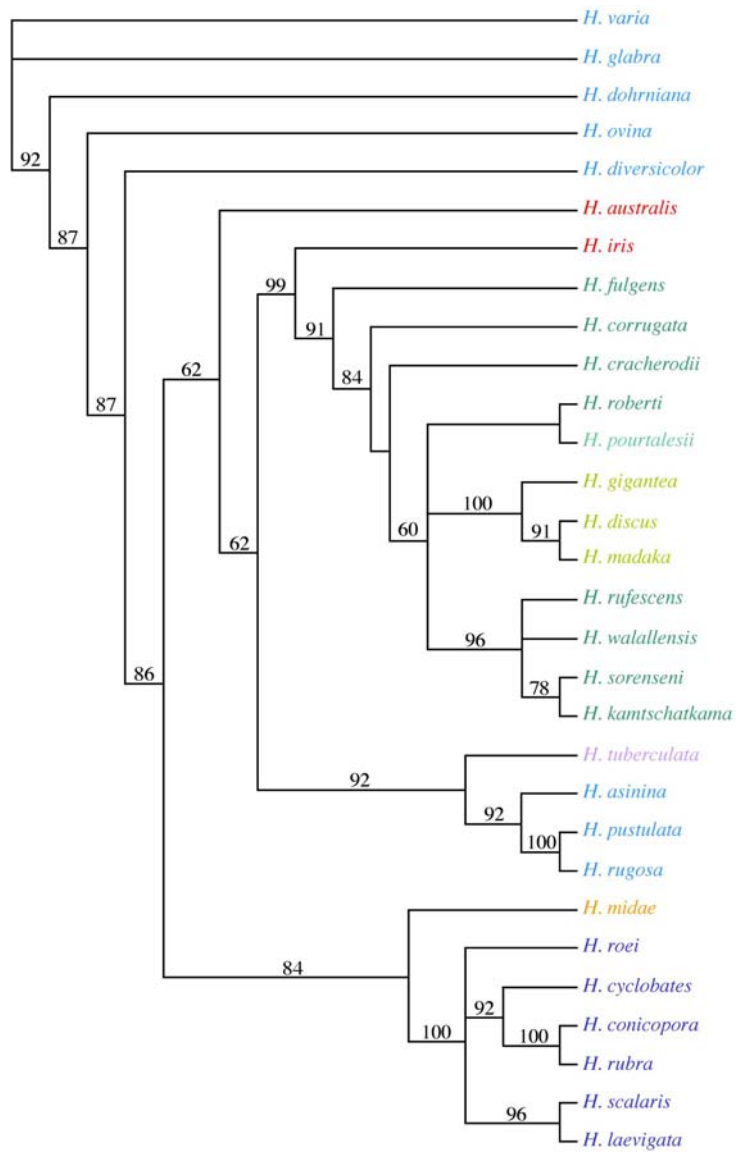


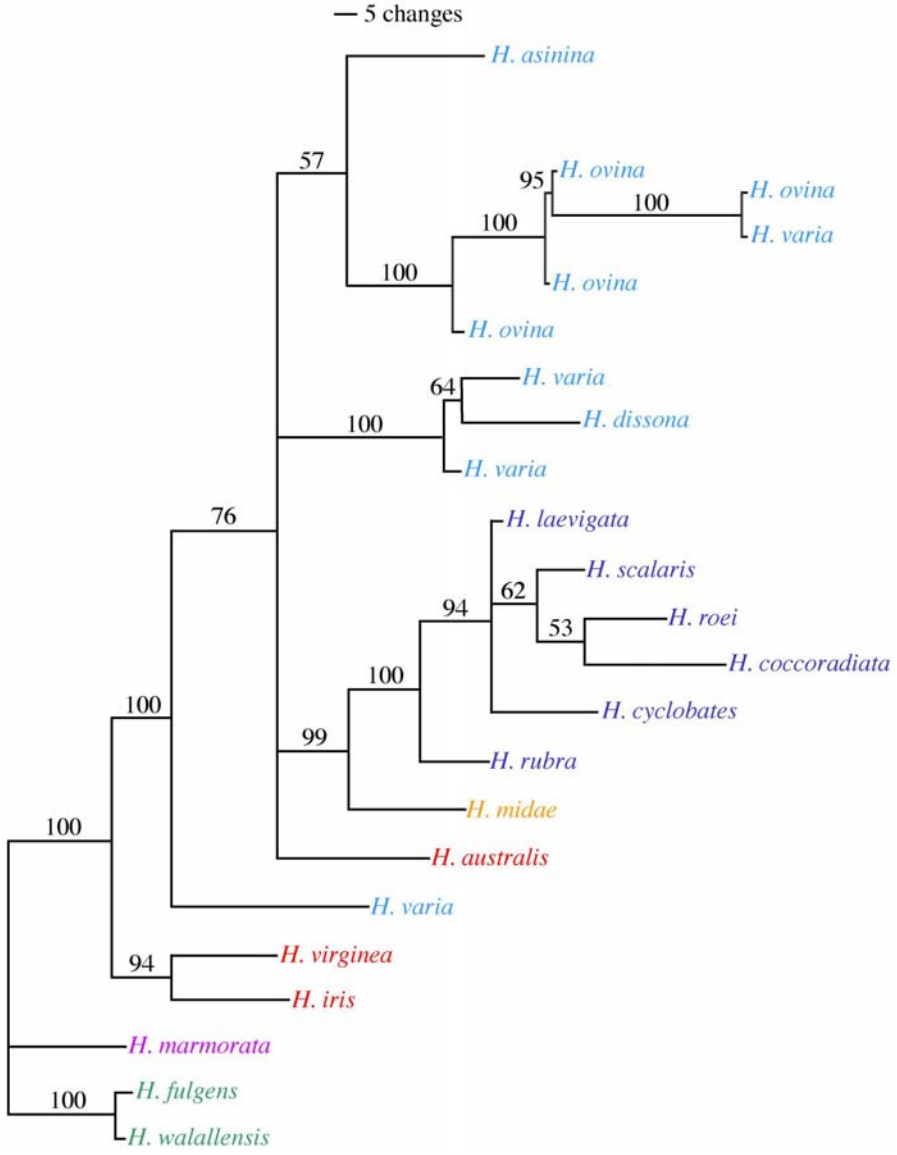
Figure 1.1: Distribution of *Haliotis* spp.. Colors represent areas where abalone species have been found and refer to regions discussed in the text, Figure 1.2, Table 1.1, and Table 1.2. The distributions are based on Geiger (2000).


A
H. rufescens
H. sorenseni
H. kamtschatkana
H. walallensis
H. sieboldii
H. discus hannai
H. cracherodii
H. corrugata
H. fulgens
H. roei
H. scalaris
H. laevigata
H. cyclobates
H. rubra
H. midae
H. pustulata
H. t. tuberculata
H. t. coccinea
H. varia
H. ovina
H. diversicolor
H. australis
H. iris
96
99
97
100
70
96
86
79
100
100
92
B
H. rubra
H. rubra
H. rubra
H. conicopora
H. roei
H. laevigata
H. scalaris
H. cyclobates
H. australis
H. iris
H. midae
H. coccoradiata
H. discus
H. virginea
2.0
1.5
1.0
0.5
0.0
C
H. roei
H. scalaris
H. laevigata
H. rubra
H. cyclobates
H. coccoradiata
H. pustulata
H. coccinea
H. tuberculata
H. australis
H. ovina
H. varia
H. diversicolor
H. midae
H. virginea
H. iris
H. rufescens
H. sorenseni
H. kamtschatkama
H. walallensis
H. gigantea
H. discus
H. corrugata
H. cracherodii
H. fulgens

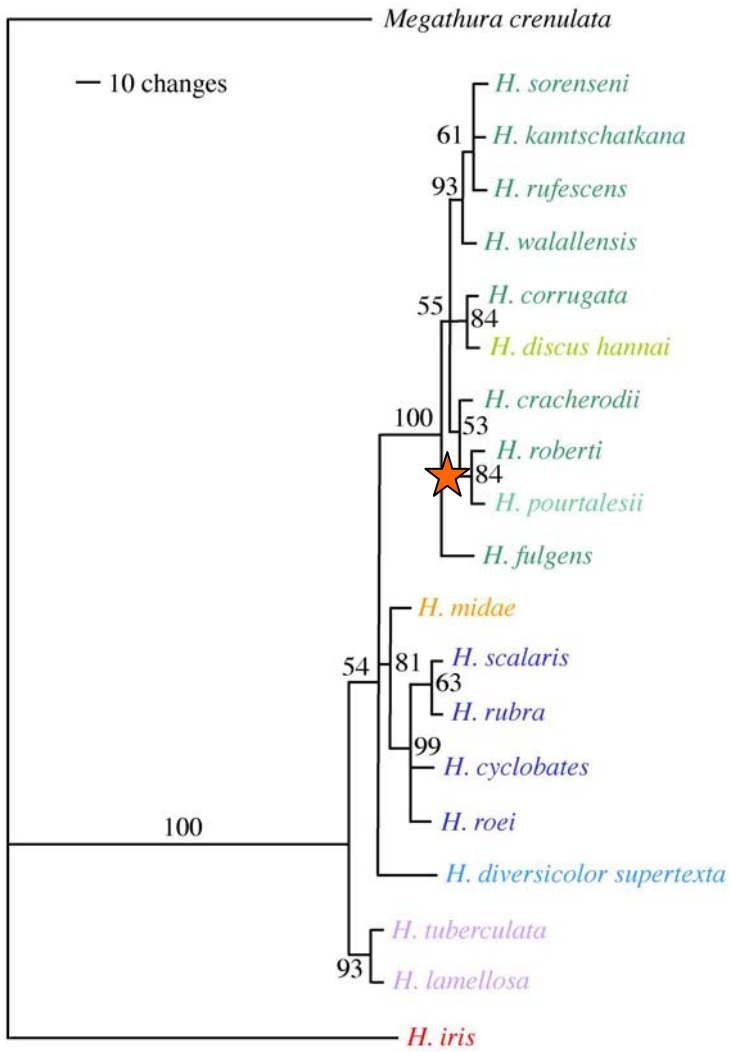
D


H. varia
H. glabra
H. dohrniana
H. ovina
H. diversicolor
H. australis
H. iris
H. fulgens
H. corrugata
H. cracherodii
H. roberti
H. pourtalesii
H. gigantea
H. discus
H. madaka
H. rufescens
H. walallensis
H. sorenseni
H. kamtschatkama
H. tuberculata
H. asinina
H. pustulata
H. rugosa
H. midae
H. roei
H. cyclobates
H. conicopora
H. rubra
H. scalaris
H. laevigata

E


5 changes
H. asinina
H. ovina
H. ovina
H. varia
H. ovina
H. ovina
H. varia
H. dissona
H. varia
H. laevigata
H. scalaris
H. roei
H. coccoradiata
H. cyclobates
H. rubra
H. midae
H. australis
H. varia
H. virginea
H. iris
H. marmorata
H. fulgens
H. walallensis

F



G

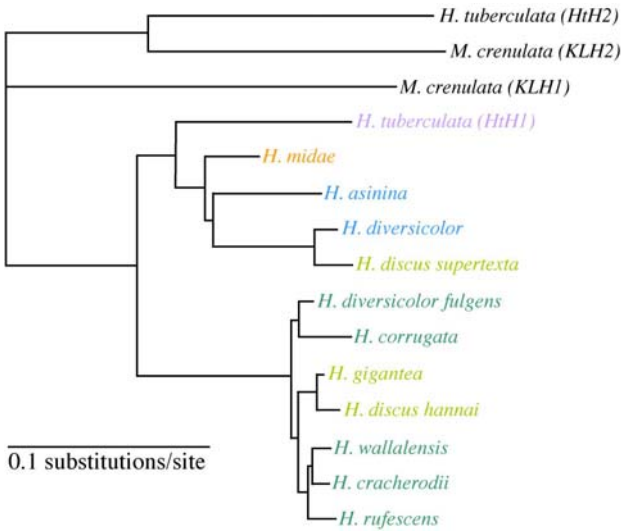


Figure 1.2: Abalone phylogenies. Colors refer to regions in Figure 1.1. All trees are unrooted except for tree F and G. Bootstrap values greater than 50% are listed when available. A) Maximum parsimony tree based on coding sequence of lysin (Lee and Vacquier 1995). B) UPGMA based on Nei's genetic distance for ten allozyme loci (Brown and Murray 1992a). C) Maximum parsimony tree combining Lee and Vacquier (1995) and Brown and Murray (1992a) in Geiger (2000). D) Consensus maximum parsimony tree based on 16S, cytochrome oxidase I, ITS, and lysin coding sequence (Estes et al. 2005). E) Modified consensus Baysian tree and posterior probabilities based on mitochondrial cytochrome oxidase II (Degnan et al. 2006). F) Rooted, consensus neighbor joining tree based on ITS1 and ITS2 (Coleman and Vacquier 2002). Star indicates Meso-American vicariance event used to date branches. G) Rooted, consensus Bayesian tree based on haemocyanin type I (Streit et al. 2006).

Table 1.1: Summary of abalone population genetics research that used nuclear data. Included in the table are nuclear data from full-length, English language articles, accessed prior to June 2008. Mitochondrial data is presented in Table 1.2. The most consistent measures reported across studies were the number of sampling sites, the distances between wild sampling sites (estimated in Google™ Earth 4.3, when possible), the mean sample size per locus for each sampling site, the mean number of alleles per locus for each sampling site, the observed and expected heterozygosities for each sampling site, and the F_{ST} value (this value was either overall F_{ST}, mean F_{ST} across loci, or an F_{ST} analogue (* indicates significance, ^{NS} indicates non-significance, ^{NA} significance was either not tested or not reported). Information is only recorded if no easily identified discrepancies regarding the values existed in the article. Conclusion is based on original authors' interpretation of the data. Regions refer to Figure 1.1.

Species	Markers	Number of samples	Distance (km)	Mean sample size	No. of loci	Mean no. of alleles	H_O	H_E	F_{ST}	Conclusion	Reference
North America											
H. rufescens	Allozymes	9 (2 groups)	20–670	116–193	4	3.25–4.00				No differentiation	(Gaffney et al. 1996)
H. rufescens	Allozymes	3	210–740	30–54	3	3.33–4.00			0.012^{NS}	No differentiation	(Burton and Tegner 2000)
H. rufescens	Microsatellites	2		39, 35	1	17, 12	0.343, 0.590	0.797, 0.711		Inconclusive	(Kirby et al. 1998)
H. rufescens	Microsatellites	9	40–950	24–61	5				0.002*	Differentiation	(Gruenthal et al. 2007)
H. rufescens	AFLP	5	140–950	24–60	163				0.035*	Differentiation	(Gruenthal et al. 2007)
H. cracherodii	Allozymes	7	10–310	23.7–119	3	3.00–3.67			0.039*	Differentiation	(Hamm and Burton 2000)
H. cracherodii	Allozymes	7	20–490	13.4–117.4	5	2.40–3.60	0.215–0.440	0.231–0.437	0.008*	Differentiation	(Chambers et al. 2006)
H. cracherodii	Microsatellites	11	<5–490		4				0.002^{NS}	Differentiation	(Gruenthal and Burton 2008)
H. cracherodii	AFLP	6	60–490		142				0.044*	Differentiation	(Gruenthal and Burton 2008)
H. fulgens	Allozymes	5	<5–70	18–22	7	1.7–2.0	0.054–0.195	0.085–0.287	0.046^{NS}	No Differentiation	(Zúñiga et al. 2000)
H. fulgens	Microsatellites	9	20–750	40.3–51.8	4	11.75–15.00	0.687–0.737	0.712–0.738	0.00062*	Differentiation	(Gutiérrez-Gonzalez et al. 2007)
H. corrugata	Allozymes	6	<5–40	15.4–31.8	8	1.9–2.8	0.094–0.201	0.141–0.258	0.093*	Differentiation	(del Río Portilla and González-Avilés 2001)
H. kamtschatkana	Microsatellites	32	<5–1200	35–180	8	13.9–15.1	0.73–0.79	0.91–0.93	0.002*	Differentiation	(Withler et al. 2003)
North Pacific											
H. discus hannai	Microsatellites	2	130	70	6	23, 21.8	0.721, 0.700	0.793, 0.793	0.004^{NS}	No differentiation	(Li et al. 2004)
H. discus hannai	Microsatellites	2	250	46, 51	9	8.67, 9.11	0.652, 0.601	0.664, 0.625	0.048*	Differentiation	(Sekino et al. 2005)
H. discus hannai	Microsatellites	5	50–430	96	8	12.8–14.1	0.643–0.684	0.666–0.674		Differentiation	(Hara and Sekino 2005)
H. discus discus	Microsatellites	5	50–1080	68.8	8	11.1–12.5	0.639–0.660	0.664–0.625		Differentiation	(Hara and Sekino 2005)
Indo-Pacific											
H. asinina	18s rDNA RFLP (3)	3	200–2170	12–28	900 bp					No differentiation	(Klinbunga et al. 2003)
H. asinina	RAPD	3	200–2170	19–28	113					Differentiation	(Tang et al. 2005)
H. asinina	Microsatellites	3	200–2170	23, 28	3	11.3–6.3	0.437, 0.607	0.683, 0.707		Differentiation	(Tang et al. 2005)
H. ovina	18s rDNA RFLP (3)	4	120–2510	11–24	900 bp					Differentiation	(Klinbunga et al. 2003)
H. varia	18s rDNA RFLP (3)	2	140	2,21	900 bp					Inconclusive	(Klinbunga et al. 2003)
Southern Africa											
H. midae	Allozymes	5	210–1510	47.8–50.2	7	3.8–4.4	0.264–0.306	0.301–0.326	0.031^{NA}	Differentiation	(Evans et al. 2004b)
H. midae	Microsatellites	6	50–790	11–48.7	3	3.67–10.7	0.333–0.471	0.490–0.567		Differentiation	(Evans et al. 2004b)[D]
H. midae	Microsatellites	6 (2 groups)	50–790	48.7, 51	3	10.7, 6.7	0.418, 0.396	0.567, 0.490		Differentiation	(Evans et al. 2004a) [D]

Australia											
H. rubra	Allozymes	17	<5–2580	47–126	12	1.67–2.87	0.099–0.150	0.101–0.163	0.015[NA]	Differentiation	(Brown 1991)[A]
H. rubra	Allozymes	15	5–1790		12				0.016[NA]	Differentiation	(Brown and Murray 1992b)[A]
H. rubra	Minisatellites	10 (4 groups)	10–880	10–60	2	4-7				Inconclusive	(Huang et al. 1997)[B]
H. rubra	Minisatellites	10	10–880		2				0.001 [NS]	Differentiation	(Huang et al. 2000) [B]
H. rubra	Microsatellites	10	10–880		3				0.067*	Differentiation	(Huang et al. 2000)
H. rubra	RAPDs	10	10–880		84				0.074*	Differentiation	(Huang et al. 2000)
H. rubra	Microsatellites	5	90–760	49.2–94.6	5	14.2–21	0.601–0.648	0.738–0.773	0.0034*	Differentiation	(Conod et al. 2002)
H. rubra	Microsatellites	7 (1 group)		576.8	5	26.8	0.578	0.648		No differentiation	(Evans et al. 2004a)
H. rubra	Microsatellites	18	<5–50	30	3	4.3–7.7	0.333–0.578	0.414–0.635	0.021*	Differentiation	(Temby et al. 2007)
H. laevigata	Allozymes	8	10–1620		13				0.014[NA]	Differentiation	(Brown and Murray 1992b)
H. roei	Allozymes	10	10–3000	44.9–56.5	8	1.5–3.13			0.0087*	Differentiation	(Hancock 2000)
New Zealand											
H. iris	Allozymes	5	340–1950		2	2				No differentiation	(Dollimore 1977)
H. iris	Allozymes	3	150–810	58–90.5	2	3.5–6.5				No differentiation	(Frusin 1982)
H. iris	Allozymes	2	780	76.5, 107	2	2.5, 3	0.380, 0.285	0.473, 0.385		Inconclusive	(Smith and Conroy 1992)
H. iris	Microsatellites	4	870–1700	20–26	6	14–22.5			0.048*	Differentiation	(Smith and McVeagh 2006)
Hatchery											
H. discus hannai	Microsatellites	3		80	6	3.5–8.5	0.517–0.621	0.559–0.715		Differentiation	(Li et al. 2004)
H. discus hannai	Microsatellites	5		40–47	7	8.0–9.4	0.517–0.596	0.754–0.787		Differentiation	(Li et al. 2007a)
H. iris	Allozymes	2		53.5, 74.5	1	2, 2	0.291–0.141	0.249–0.131		Inconclusive	(Smith and Conroy 1992)
H. fulgens	Microsatellites	2		127, 50	2	17, 13.5	0.863, 0.676	0.836–0.867		Inconclusive	(Gutiérrez-Gonzalez and Perez-Enriquez 2005)
H. asinina	RAPDs	3		15–30						Inconclusive	(Tang et al. 2005)
H. asinina	Microsatellites	2		15, 15	3	7, 6	0.54, 0.53	0.74, 0.74		Inconclusive	(Tang et al. 2005)
H. midae	Microsatellites	2		127, 52	3	7.7, 4.7	0.471, 0.426	0.573, 0.508		Differentiation	(Evans et al. 2004a)
H. rubra	Microsatellites	4 (1 group)		255	5	16	0.574	0.603		Inconclusive	(Evans et al. 2004a)
H. rufescens	Allozymes	4		43–156	4	2.3–2.5				Differentiation	(Gaffney et al. 1996)
H. varia	18s rDNA RFLP (4)	3		15–20	900 bp					Inconclusive	(Klinbunga et al. 2003)

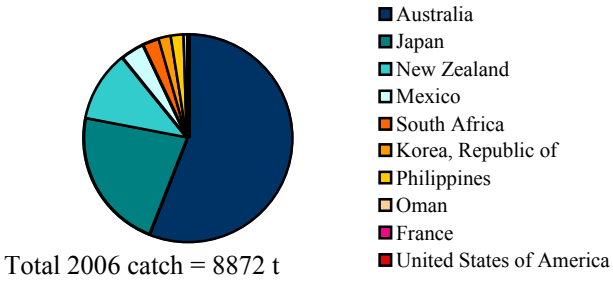
[A] Overlapping data
[B] Overlapping data
[C] Pooled offspring data
[D] Overlapping data

Table 1.2: Summary of abalone population genetics research that used mitochondrial data. Included in the table are mitochondrial data from full-length, English language articles. Nuclear data is presented in Table 1.1. The most consistent measures reported across studies were the number of sampling sites, the distances between wild sampling sites (estimated in Google™ Earth 4.3, when possible), the sample size for each sampling site, the length of the fragment amplified, the number of haplotypes per sampling site, haplotype and nucleotide diversities per sampling site. Information is only recorded if no easily identified discrepancies regarding the values existed in the article. Conclusions were based on the original authors' interpretation of the data. Regions refer to Figure 1.1.

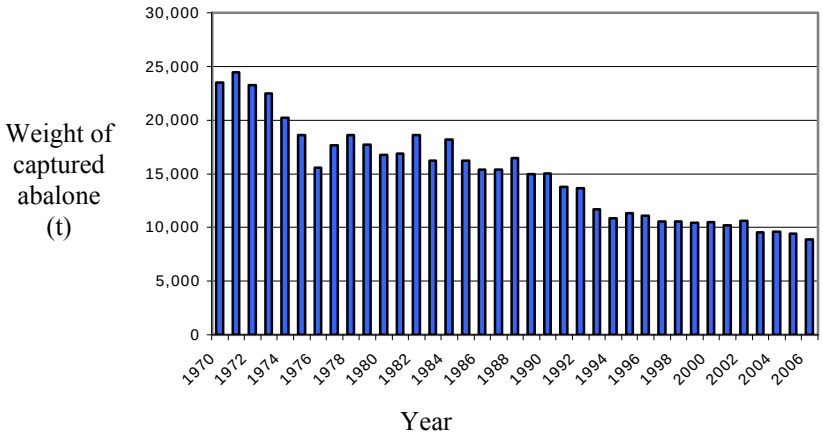
Species	Markers	No. of samples	Distance (km)	Sample sizes	Length (bp)	No. of haplotypes	Haplotype diversity	Nucleotide diversity	Conclusion	Reference
North America										
H. rufescens	COI	3	210–740	20–21	484	5–7	0.81–0.87	0.0042–0.0005	No differentiation	(Burton and Tegner 2000)
H. rufescens	COI	9	40–950	29–36	580	7–14			No differentiation	(Gruenthal et al. 2007)
H. cracherodii	COI	5	10–310	8–12	382	3–6			No differentiation	(Hamm and Burton 2000)
H. cracherodii	COI	11	<5–490	11–27	580	32 [B]	0.630 [B]	0.002 [B]	Differentiation	(Gruenthal and Burton 2008)
Indo-Pacific										
H. diversicolor	Genome RFLP (2 [A])	5	10–240	12	173300–19440	1–2			Inconclusive	(Jiang et al. 1995)
H. asinina	16S rDNA RFLP (4 [A])	3	110–2150	12–28	580	1–2			No differentiation	(Klinbunga et al. 2003)
H. asinina	COII	16	10–7010	2–20	482	1–9	0.0000–0.9722	0.000000–0.017404	Differentiation	(Imron et al. 2007)
H. ovina	16S rDNA RFLP (4 [A])	4	120–2510	11–24	580	1–2			Differentiation	(Klinbunga et al. 2003)
H. varia	16S rDNA RFLP (4 [A])	2	140	2, 21	580	2, 5			Inconclusive	(Klinbunga et al. 2003)
Southern Africa										
H. midae	NADH1-16S and COII-NADH3 RFLP (8 [A])	16	10–1120	18–39	3300	3–10	0.6377–0.8918	0.0000049–0.000088	Differentiation	(Evans et al. 2004b)
Australia										
H. rubra	ND3-COIII RFLP (6 [A])	5	90–760	36–39	1500	8–14	0.63–0.82		Differentiation	(Conod et al. 2002)
New Zealand										
H. virginea	16S	4	1000–2400	3–9	500	2–8			Differentiation	(Clarke 2001)
H. iris	ATP8–ATP6	4	1700	8–11	631	3–6			Differentiation	(Smith and McVeagh 2006)
Hatchery										
H. asinina	16S rDNA RFLP (4 [A])	3		15-20	580	1-2			Inconclusive	(Klinbunga et al. 2003)

[A] Number of restriction enzymes used
[B] Values are for all samples

A



B



C

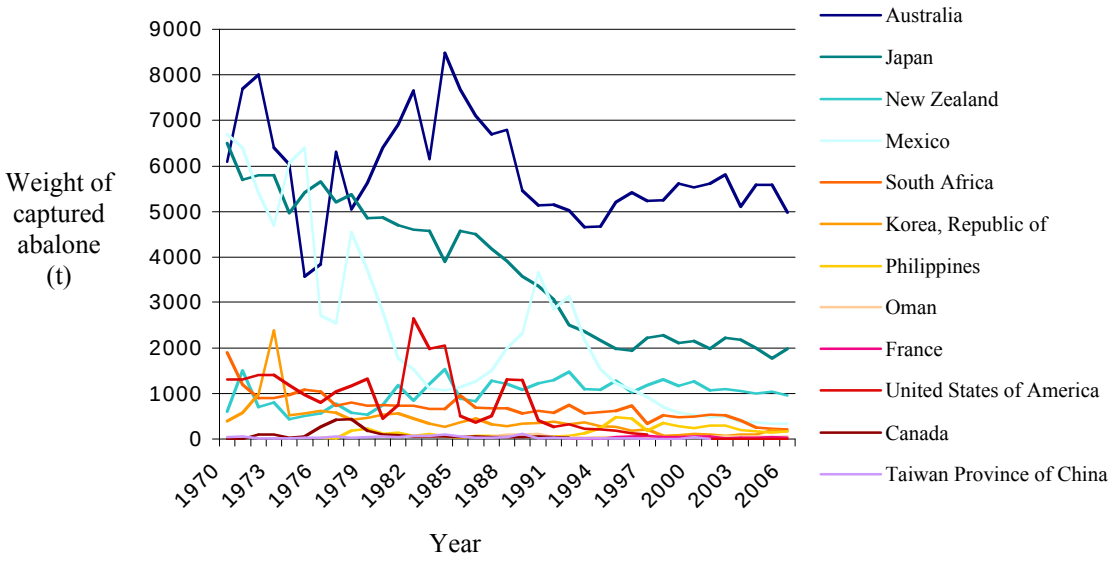


Figure 1.3: World abalone catch. A) The 2006 world catch of abalone species (*Haliotis* spp.) partitioned according to fishing countries. B) Annual total world catch of abalone species from 1970–2006. C) Annual total catch for fishing countries from 1970–2006. Channel Islands, Solomon Islands, and Ireland never fished over 68 t and were excluded for clarity in graph C. Data is from FAO (2000).

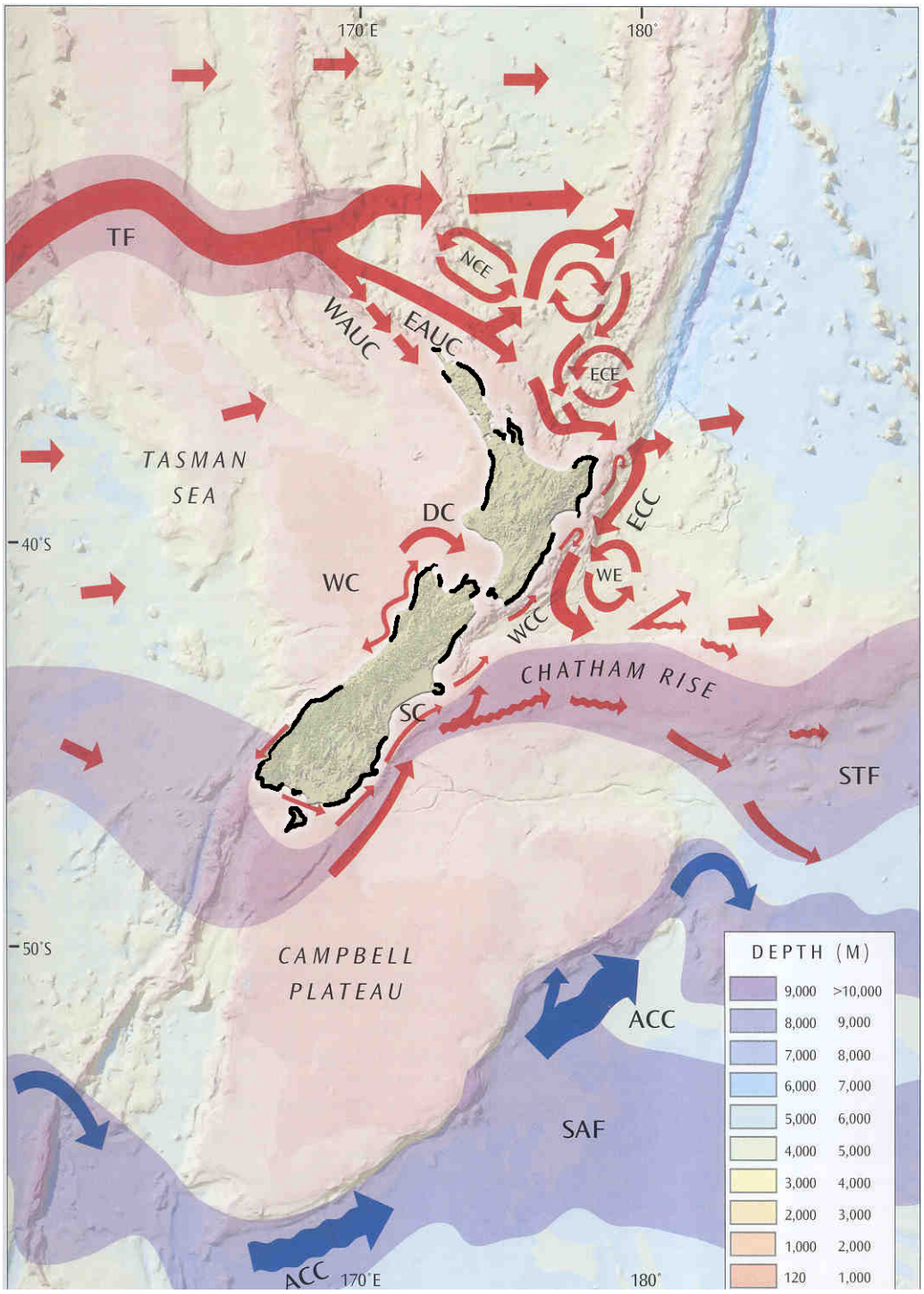


Figure 1.4: Ocean currents and rocky reefs around New Zealand. Black border around New Zealand represents rocky coast off of which rocky reefs are often found. Abbreviations: *TF*, Tasman Front; *STF*, Subtropical Front; *SAF*, Subantarctic Front; *EAUC*, East Auckland Current; *WAUC*, West Auckland Current; *ECC*, East Cape Current; *DC*, D'Urville Current; *WC*, Westland Current; *WCC*, Wairarapa Coastal Current; *SC* Southland Current; *ACC*, Antarctic Circumpolar Current; *NCE*, North Cape Eddy; *ECE*, East Cape Eddy; *WE*, Wairarapa Eddy. Modified from Laing (2003 pp. 22 and 25).

Table 2.1: Summary of New Zealand marine invertebrate population genetic research. This table is similar to Table 5.1 in Goldstien (2005) and Table 4.1 in Veale (2007).

Taxon	Markers	No. of loci[A]	No. of samples[B]	Sample sizes[C]	Sampling range[D]	Structure	Reference
Jasus edwardsii	allozymes	1	3	36–54	NE, CN, ST	No structure	(Smith et al. 1980)
Jasus edwardsii	mtDNA RFLP	6 (15966 bp)	2	9, 10	NE, SE, AUS	No structure	(Ovenden et al. 1992)
Pinnotheres atrinicola	allozymes	15 (23)	7	20–70	N, NW, CN	[N][NW, CN], clinal within the N	(Stevens 1991)
Paracorophium lucasi	allozymes	9 (10)	9	22–38	N, NW, CN	[N][NW, CN]	(Schnabel et al. 2000)
Paracorophium lucasi	allozymes	12	18	11–41	N, NE, NW, CN, CS	[N][NE][N, NW, CN, CS]	(Stevens and Hogg 2004)
Paracorophium excavatum	allozymes	9 (10)	4	13.4–36.6	N, CS, SE	[N][CS][SE]	(Schnabel et al. 2000)
Paracorophium excavatum	allozymes	12	21	24–80	N, NE, CN, CS, SE, SW, S, CH	[N][NE, CN, CS, SE, SW, S][CS][CH]	(Stevens and Hogg 2004)
Errina novaezelandiae	allozymes	5 (9)	9	7.4–39.2	F	Structure	(Miller et al. 2004)
Actinia tenebrosa	microsatellites	4	26	3–24	N, NE, NW, CN, CS, SE, SW, AUS	Isolation by distance	(Veale 2007)
Evechinus chloroticus	allozymes	4 (5)	6	18–68	N, NE, SE, ST, F	[N, NE, SE, ST] [F]	(Mladenov et al. 1997)
Evechinus chloroticus	microsatellites	6	8	30–43	N, NW, SW, SE ST, F	[F] [F, ST] [N, NW] [SE, SW]*	(Perrin 2002)
Evechinus chloroticus	microsatellites	6	29	28–40	F	[inner fiord][outer fiord]	(Perrin 2002)
Amphipholis squamata	16S	486 bp	16	4–17	N, NW, CN, CS, SE, S, ST, F	[N, NW, CN, CS, SE][SE. S, ST, F]	(Sponer and Roy 2002)
Coscinasterias muricata	allozymes	4	16	24-76	NW, SE, ST, F	[NW][SE, ST][F]	(Sköld et al. 2003)
Coscinasterias muricata	D-loop	318 bp	17	17–32	NW, CS, ST, F	Isolation by distance within fiords [ST][CS][NW][F]	(Perrin 2002) (Perrin et al. 2004)
Patriella regularis	control region	822 bp	19	4–7	N, NW, NE, CN, CS, SW, SE, ST, F, AUS	[N, NW, NE, CN, CS][SW, SE, ST, F] within NI: [N, NE][NW]	(Waters and Roy 2004)
Patriella regularis	control region	835 bp	22	4–23	N, NW, NE, CN, CS, SW, SE, S, ST, F	[N, NW, NE, CN, CS] [SW, SE, S, ST, F]	(Ayers and Waters 2005)
Crassostrea giga	allozymes	8 (9)	2	28–52	NE, NW, JAP	No structure	(Smith et al. 1986)
Perna canaliculus	allozymes	8 (10)	6	68–104	N, NW, NE, CN, S, SE	[N, NW] [NE, CN, S, SE]	(Smith 1988)
Perna canaliculus	allozymes	7 (11)	9 (1)	4–141	N, NW, NE, CN, CS, SE, ST	Isolation by distance	(Gardner et al. 1996)
Perna canaliculus	allozymes	7	31 (4)	20–39	N, NW, NE, CN, CS, S, SW, SE, ST, F	No structure	(Apte and Gardner 2001)
Perna canaliculus	NADHIV	391 bp	18 (4)	26	N, NW, NE, CN, CS, S, SW, SE, ST, F	[N, NW, NE, CN, CS] [S, SW, SE, ST, F]	(Apte and Gardner 2002)
Perna canaliculus	RAPDs (3)	21 bands	14 (5)	20–31	N, NW, NE, CS, S, SW, SE, ST, F	[N, NW, NE, CS] [S, SW, SE, ST, F]*	(Star et al. 2003)
Paphies subtriangulata	allozymes	4 (9)	13	16–114	N, NW, NE, CN, CS, ST, CI	[N, NW, NE, ST] [CN, CS] [CI]*	(Smith et al. 1989)
Cellena ornate	cytb	328 bp	31	5–15	N, NW, NE, CN, CS, SW, SE, ST	[N, NW, NE, CN, CS] [SW, SE, ST]	(Goldstien et al. 2006b)
Cellana radians	cytb	328 bp	31	4–24	N, NW, NE, CN, CS, SW, SE, ST	[N, NW, NE, CN, CS, SE] [SW, SE, ST]	(Goldstien et al. 2006b)
Cellana flava	cytb	359 bp	8	8–20	N, NE, SE	[N, NE] [SE]	(Goldstien et al. 2006b)
Sypharochiton pelliserpentis	COI (RFLP)	706 bp (4)	28	3–18 (3–22)	N, NW, NE, CN, CS, SW, SE, S, ST	[N, NW, NE, CN, CS] [SW, SE, S, ST]	(Veale 2007)
Nerita atramentosa	COI	1107 bp	10	2–6	N, NW, AUS, Easter Island	No structure	(Waters et al. 2005)
Terebratella sanguinea	allozymes	4	10	11–52	F, ST	No structure	(Ostrow et al. 2001)
Haliotis virginea	16S	500 bp	4	2–8	N, CN, CI, CaI	[N, CN] [CI] [CaI]	(Clarke 2001)
Haliotis iris	allozymes	2	5		N, NE, CN, SE	No Structure	(Dollimore 1977)
Haliotis iris	allozymes	2	3	60–248	CN, SE, CI	[CN, SE][CI]	(Frusin 1982)
Haliotis iris	allozymes	2	2	75–108	CN, CI	Inconclusive	(Smith and Conroy 1992)
Haliotis iris	microsatellites ATP8–ATP6	6 631 bp	4 4	20–26 8–11	N, NW, ST, CI N, CN, ST, CI	[N][NW][ST][CI] [N, NW, ST][CI]	(Smith and McVeagh 2006)

[A] For sequence data, the number of base pairs examined is given. For RFLPs, the number of restriction enzymes used is listed in parentheses. For RAPDs, the number of primers used is listed in parentheses. For allozymes, the number of polymorphic loci (followed by the number of loci examined in parentheses) is given.
[B] The number of samples includes only NZ wild samples and not overseas samples. Numbers in parentheses indicate number of cultured samples.
[C] Sample sizes are either the number of samples collected, the number successfully screened, the average across samples, or the average number successfully screened across loci.
[D] N: samples collected between Cape Reinga and East Cape, NE: the remaining North Island east coast, NW: North Island west coast, CN: North Island greater Cook Strait region, CS: South Island greater Cook Strait region, SE: South Island east coast, S: South Island south coast, SW: South Island west coast, F: Fiordland, ST: Stewart Island, CI: Chatham Islands, CaI: Campbell Island.

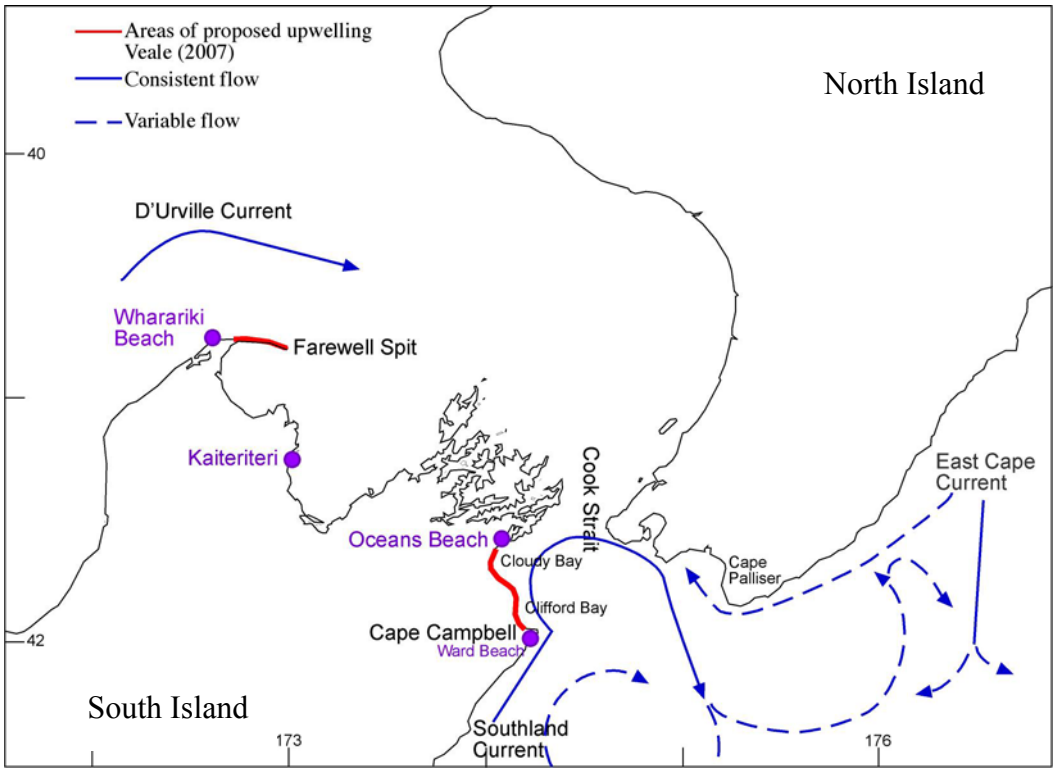
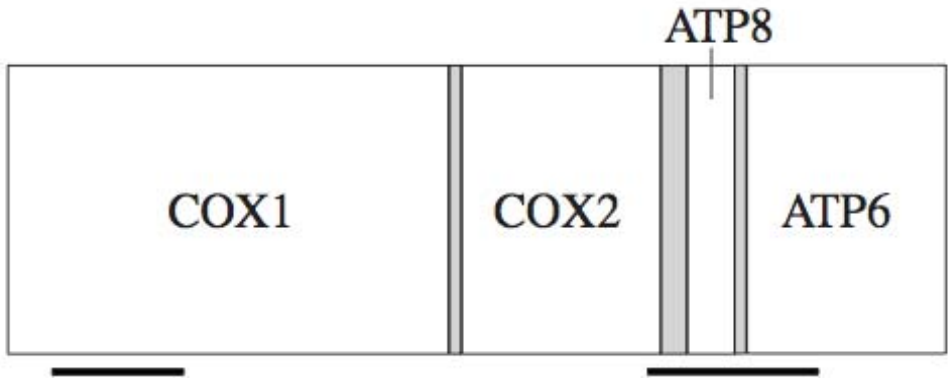


Figure 2.1: Cook Strait region. Shown are areas of upwelling proposed by Veale (2007), locations mentioned in text, and currents around the Cook Strait region. The flow and direction of the East Cape and Southland Currents are from Figure 4 in Barnes (1985).



mtCOI_F2:	TTTAGGGGACGACCAACTGTA	COIIcons-F:	ATGCTCAGAAATTTGCGGAACTAAC
mtCOI_R2:	TACGGTCGGTTAGGAGCATT	H22-R1:	TATGGCGAGGTGGCTGGTATTTC

Figure 2.2: Amplified fragments of *H. iris* mitochondrial DNA. Primer sequences and regions amplified in the *H. iris* mitochondrial genome based on the *H. rubra* mitochondrial genome (ACCN: NC_005940). Segments are drawn to scale based on Maynard et al. (2005).

Location	N
AHU	22
CBL	14
CCB	20
CRW	19
DBL	15
DSD	14
EAI	24
GLN	20
GOB	19
IHM	20
JCH	20
MAT	21
MTB	22
NPT	20
OCH	13
OLB	20
OPT	20
PHD	20
SPB	21
STR	15
TCL	20
TIM	23
TSK	20
WLG	20
WST	15

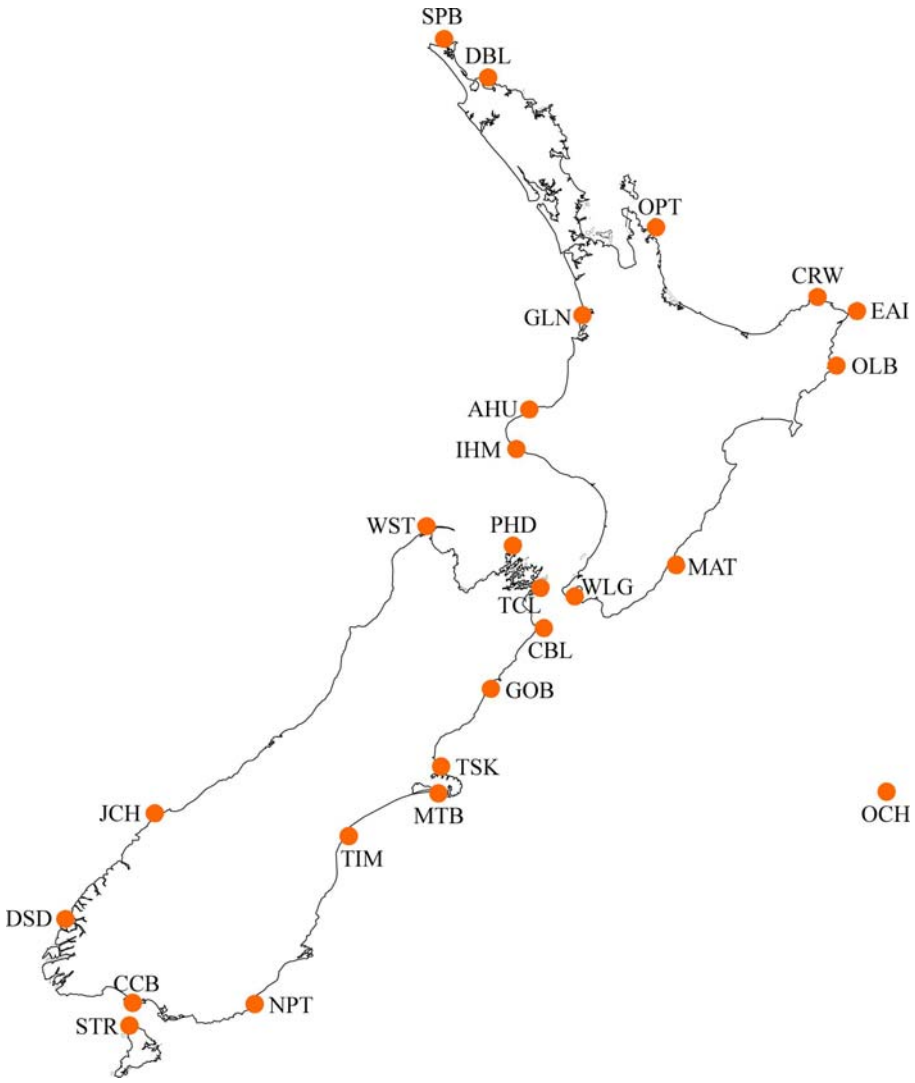


Figure 2.3: Locations and sample sizes for *H. iris* collected around New Zealand. Location of the Chatham Islands (OCH) is not to scale and is located about 946.7 km from TSK and 937.4 km from MTB. Further details about collections are given in Appendix 1.

Table 2.2: A priori population genetic structures. Groups reflected potential genetic split between Chatham Islands and mainland samples (Smith and McVeagh 2005) and between northern and southern samples (Apte and Gardner 2002; Star et al. 2003; Waters and Roy 2004; Ayers and Waters 2005; Goldstien et al. 2006b; Veale 2007). Samples refer to locations shown in Figure 2.3.

Groups	Samples
Chatham Islands North and South Island	[OCH] [SPB, DBL, OPT, CRW, EAI, OLB, MAT, WLG, IHM, AHU, GLN, WST, PHD, TCL, CBL, GOB, TSK, MTB, TIM, NPT, STR, CCB, DSD, JCH]
Chatham Islands North Island South Island	[OCH] [SPB, DBL, OPT, CRW, EAI, OLB, MAT, WLG, IHM, AHU, GLN] [WST, PHD, TCL, CBL, GOB, TSK, MTB, TIM, NPT, STR, CCB, DSD, JCH]
Chatham Islands North Island and north of South Island Remaining South Island	[OCH] [SPB, DBL, OPT, CRW, EAI, OLB, MAT, WLG, IHM, AHU, GLN, PHD, TCL] [WST, CBL, GOB, TSK, MTB, TIM, NPT, STR, CCB, DSD, JCH]
Excluding Chatham Islands	
North Island South Island	[SPB, DBL, OPT, CRW, EAI, OLB, MAT, WLG, IHM, AHU, GLN] [WST, PHD, TCL, CBL, GOB, TSK, MTB, TIM, NPT, STR, CCB, DSD, JCH]
North Island and north of South Island Remaining South Island	[SPB, DBL, OPT, CRW, EAI, OLB, MAT, WLG, IHM, AHU, GLN, PHD, TCL] [WST, CBL, GOB, TSK, MTB, TIM, NPT, STR, CCB, DSD, JCH]

Table 2.3: Polymorphism data and neutrality results for mtCOI, ATPase8–ATPase6, and concatenated mitochondrial fragments. Individuals from all sampling locations were grouped as a single sample. Results were generated in Arlequin 3.1 (Excoffier et al. 2005).

Region	N	bp	Number of polymorphic sites	Number of haplotypes	Haplotype diversity $h \pm$ S.D.	Nucleotide diversity $\pi \pm$ S.D.	Tajima's D (p-value)	Fu's F_s (p-value)
mtCOI	477	459	30	38	0.5472±0.0262	0.002457±0.001772	-1.96118 (0.002)	-28.01814 (0.000)
ATPase8–ATPase6	477	596	89	98	0.8760±0.0085	0.004883±0.002840	-2.23974 (0.000)	-25.86168 (0.000)
Concatenated	477	1055	119	132	0.8990±0.0081	0.003827±0.002117	-2.25067 (0.000)	-25.16002 (0.000)

Table 2.4: Polymorphism data and neutrality results for concatenated mitochondrial fragments across sampling locations. Individuals are grouped according to sampling locations (Figure 2.3). Results were generated in Arlequin 3.1 (Excoffier et al. 2005).

Location	N	Number of polymorphic sites	Number of haplotypes	Haplotype diversity $h \pm$ S.D.	Nucleotide diversity $\pi \pm$ S.D.	Tajima's D (p-value)	Fu's F_s (p-value)
AHU	22	22	17	0.9740±0.0217	0.003947±0.002276	-1.165 (0.115)	-10.420 (0.000)
CBL	14	11	5	0.8242±0.0567	0.004306±0.002529	1.230 (0.908)	2.382 (0.878)
CCB	20	14	5	0.8053±0.0497	0.004190±0.002410	0.436 (0.715)	3.407 (0.928)
CRW	19	15	8	0.8889±0.0420	0.003551±0.002094	-0.478 (0.354)	-0.301 (0.458)
DBL	15	12	21	0.9619±0.0399	0.004807±0.002771	-0.886 (0.192)	-4.726 (0.012)
DSD	14	9	4	0.5824±0.1372	0.002304±0.001490	-0.544 (0.330)	1.715 (0.831)
EAI	24	19	10	0.8768±0.0383	0.002293±0.001437	-1.905 (0.011)	-3.005 (0.056)
GLN	20	18	14	0.9632±0.0255	0.003236±0.001929	-1.233 (0.101)	-7.553 (0.000)
GOB	19	12	7	0.7836±0.0743	0.003118±0.001875	-0.156 (0.490)	0.163 (0.561)
IHM	20	26	16	0.9737±0.0250	0.005478±0.003056	-0.754 (0.245)	-7.287 (0.004)
JCH	20	17	7	0.8474±0.0421	0.004289±0.002460	-0.211 (0.459)	1.301 (0.758)
MAT	21	20	14	0.9238±0.0426	0.003226±0.001919	-1.459 (0.055)	-7.105 (0.000)
MTB	22	7	8	0.8182±0.0515	0.001323±0.000941	-0.877 (0.205)	-3.313 (0.011)
NPT	20	14	8	0.8579±0.0451	0.003910±0.002269	0.162 (0.612)	0.140 (0.552)
OCH	13	12	6	0.8205±0.0817	0.004890±0.002847	1.359 (0.935)	1.370 (0.758)
OLB	20	20	12	0.9263±0.0378	0.004934±0.002783	-0.295 (0.419)	-2.507 (0.126)
OPT	20	23	13	0.9316±0.0421	0.003956±0.002292	-1.314 (0.088)	-4.729 (0.015)
PHD	20	21	14	0.9526±0.0304	0.004849±0.002741	-0.521 (0.340)	-4.939 (0.014)
SPB	21	29	18	0.9810±0.0225	0.004897±0.002758	-1.391 (0.069)	-11.431 (0.000)
STR	15	15	6	0.8190±0.0636	0.004446±0.002586	0.063 (0.562)	1.545 (0.786)
TCL	20	20	13	0.9105±0.0538	0.003510±0.002068	-1.306 (0.088)	-5.424 (0.006)
TIM	23	14	10	0.8538±0.0486	0.001830±0.001203	-1.739 (0.022)	-4.212 (0.008)
TSK	20	14	8	0.8474±0.0512	0.002547±0.001579	-1.171 (0.115)	-1.207 (0.260)
WLG	20	31	16	0.9632±0.0328	0.004394±0.002512	-1.846 (0.018)	-8.951 (0.000)
WST	15	10	5	0.7810±0.0740	0.002602±0.001638	-0.412 (0.372)	1.031 (0.728)

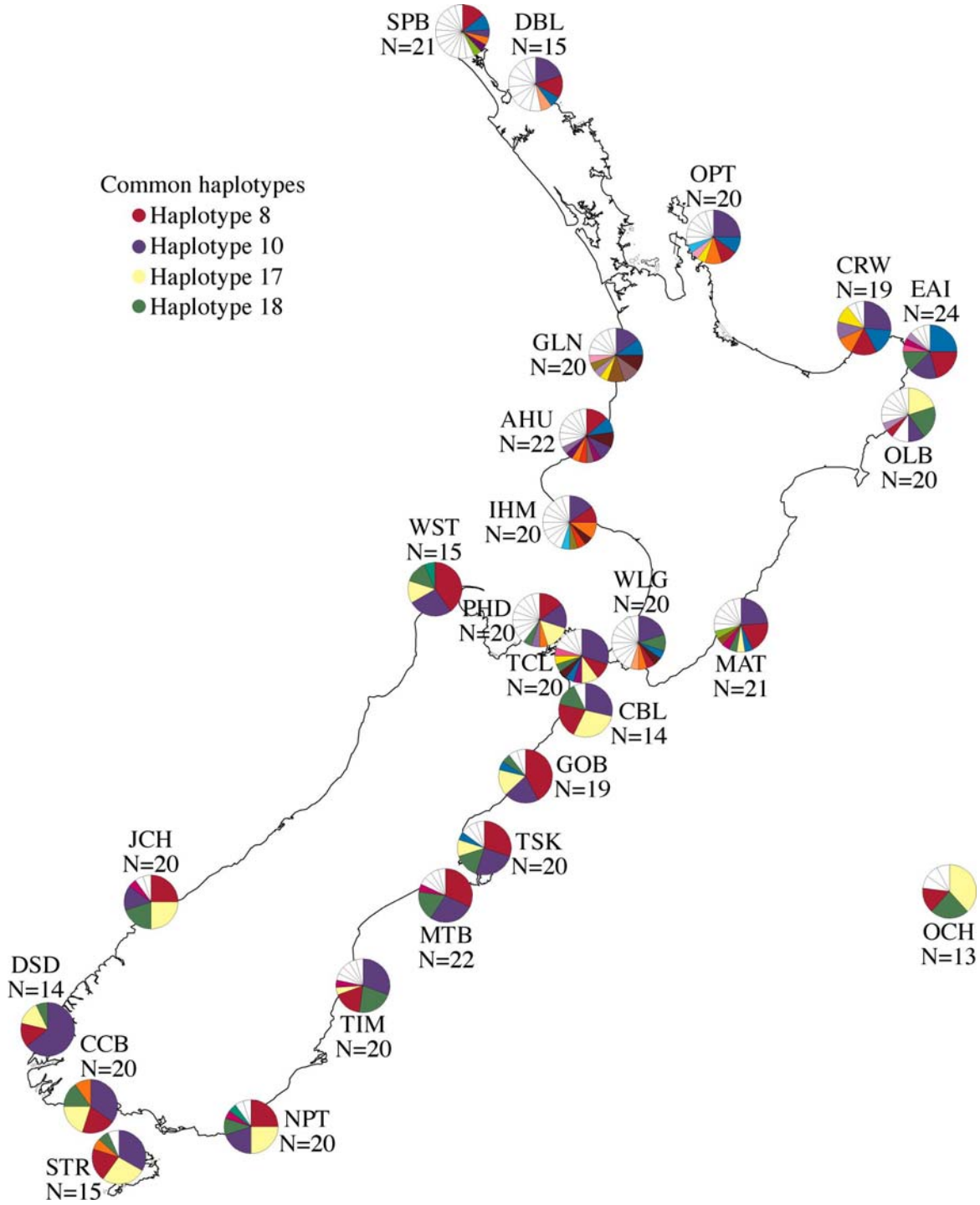


Figure 2.4: Mitochondrial haplotype frequencies and sample sizes at collection locations around New Zealand. Colored haplotypes are shared among sampling locations, while white haplotypes are private. Colors correspond to haplotypes in Figures 2.5–2.7 and 2.10. Haplotype sequences and frequencies are listed in Appendices 2 and 3, respectively.

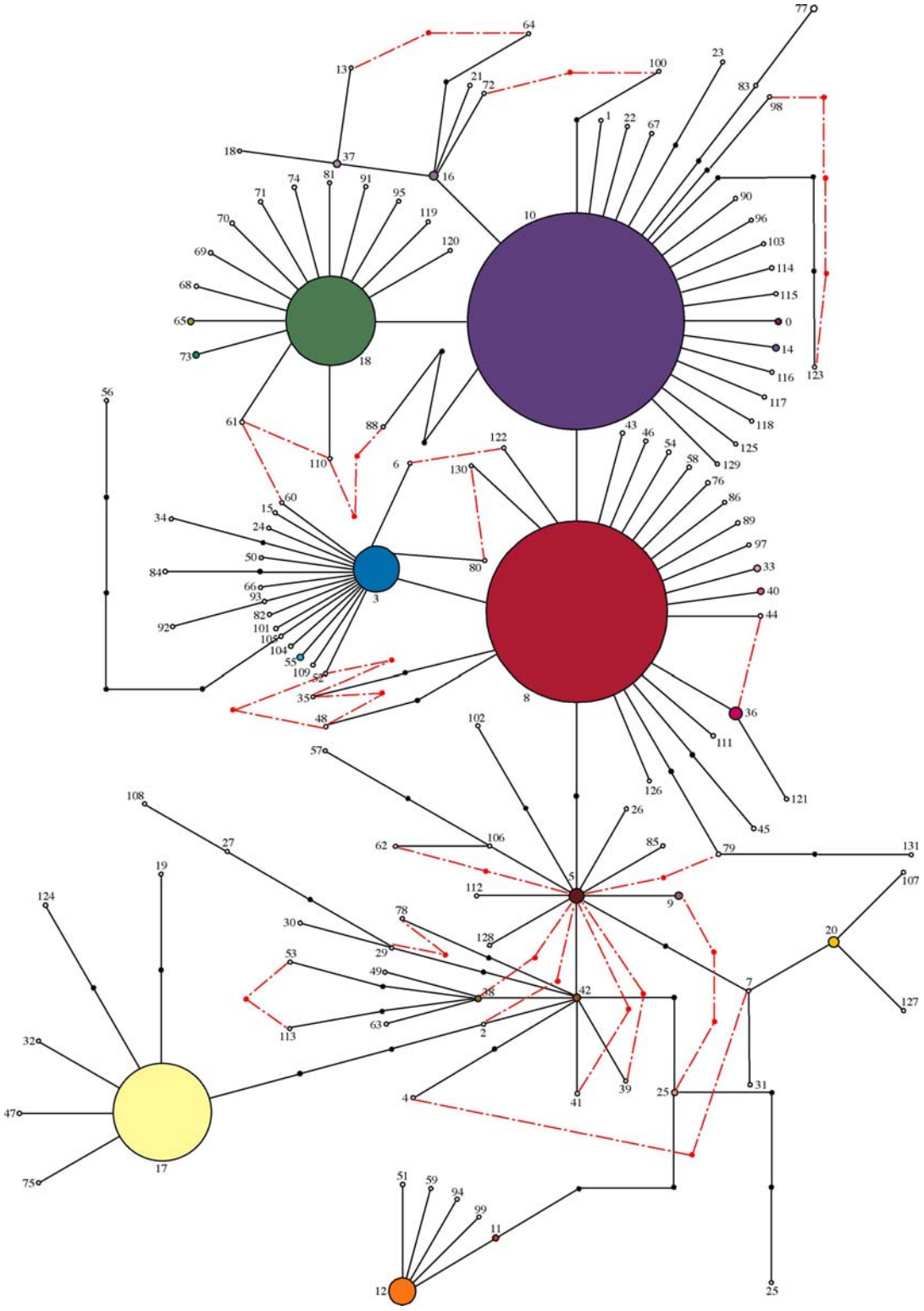


Figure 2.5: Minimum spanning network based on mitochondrial DNA. Relationships among haplotypes were inferred using a minimum spanning algorithm (Excoffier and Smouse 1994), implemented in Arlequin 3.1 (Excoffier et al. 2005). For clarity, cycles are completed with red dashed lines. Colored haplotypes are shared among sampling locations, white haplotypes are private, and black haplotypes are missing haplotypes.. Haplotypes are labeled from 0–131, and haplotype sequences are listed in Appendix 2.

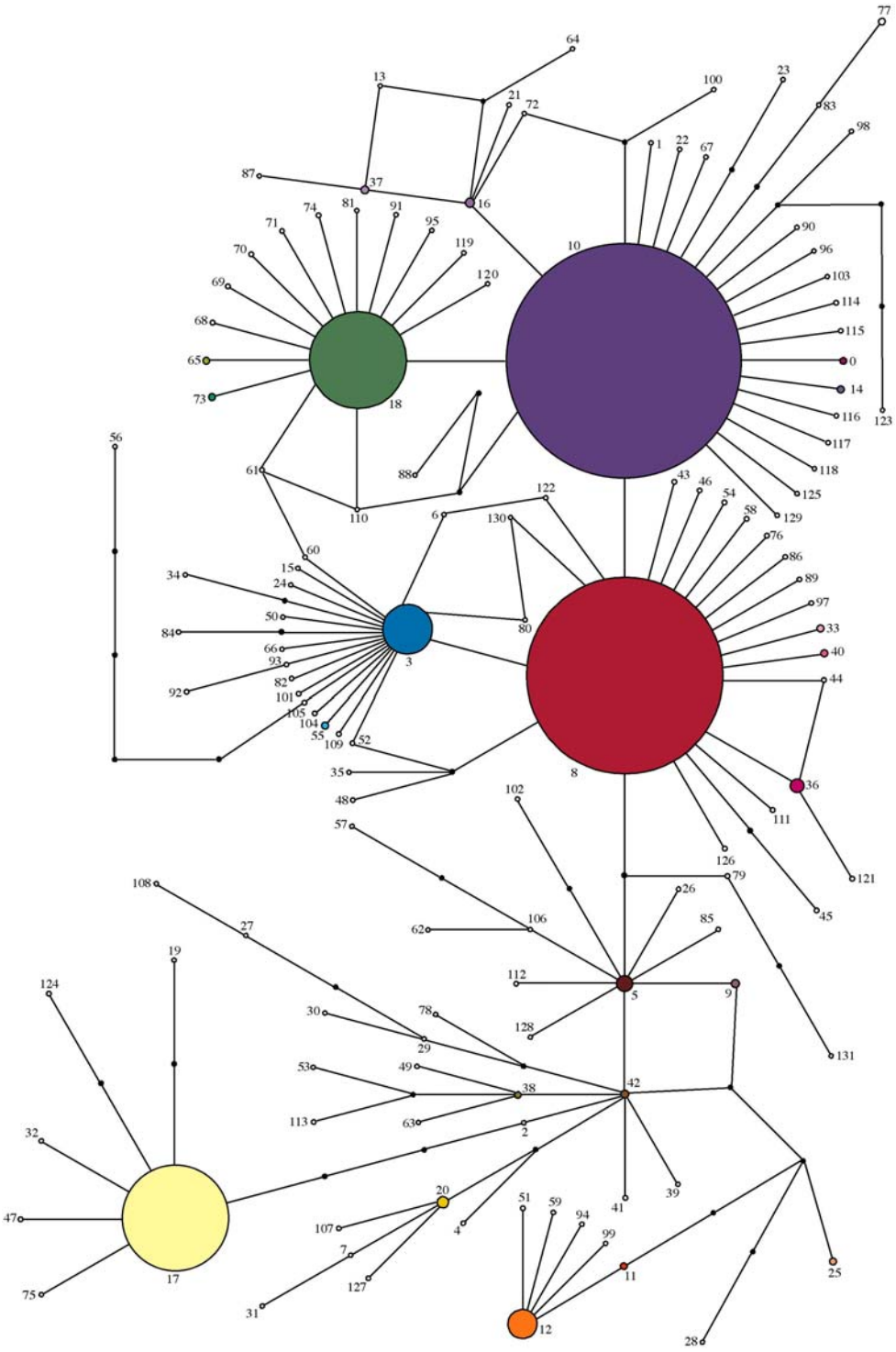


Figure 2.6: Statistical parsimony network based on mitochondrial DNA. Relationships among haplotypes were inferred using statistical parsimony (Templeton et al. 1992), implemented in TCS (Clement et al. 2000). Colored haplotypes are shared among sampling locations, white haplotypes are private, and black haplotypes are missing haplotypes. Haplotypes are labeled from 0–131, and haplotype sequences are listed in Appendix 2.

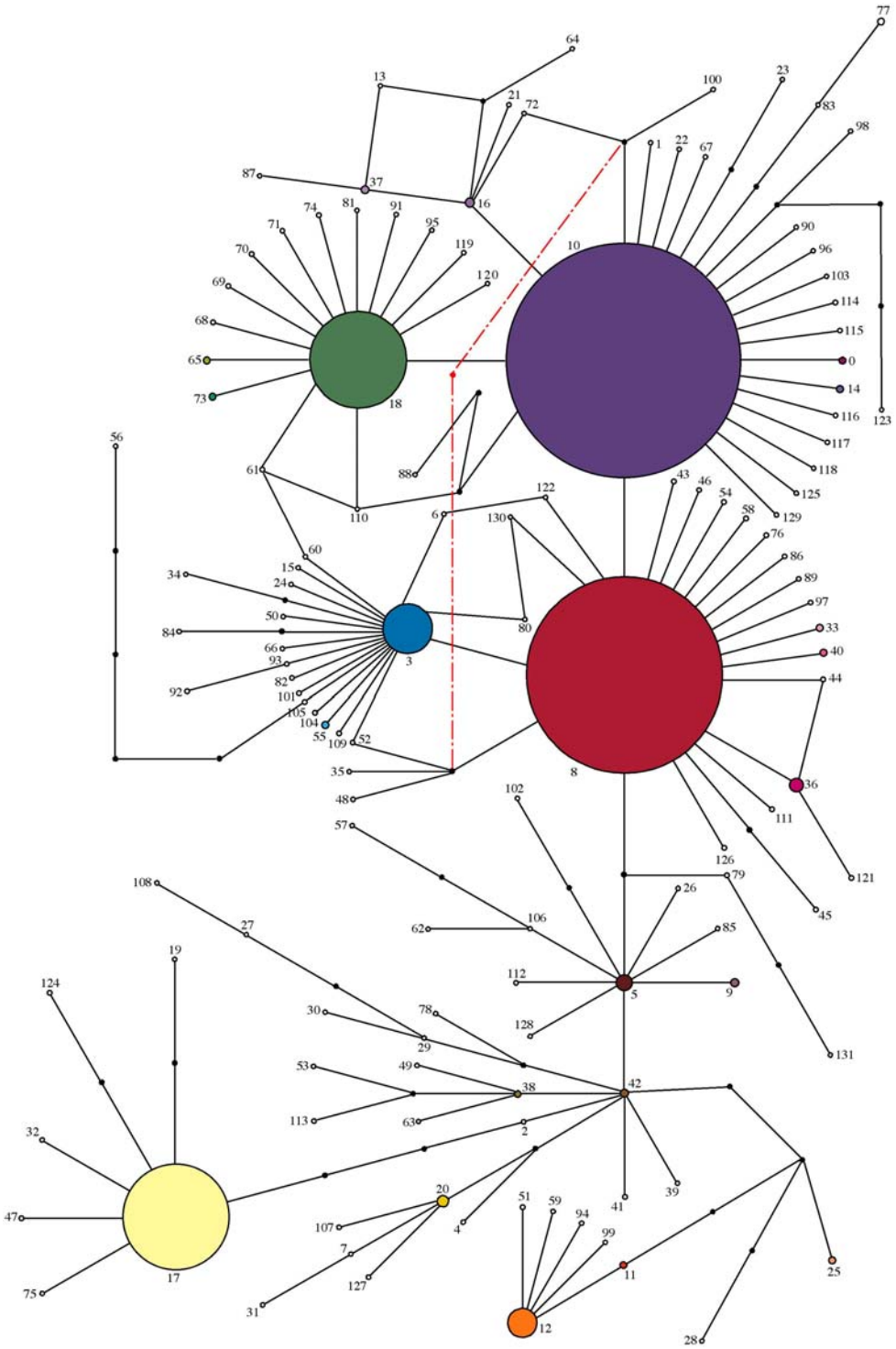


Figure 2.7: Median joining network based on mitochondrial DNA. Relationships among haplotypes were inferred using a median joining algorithm (Bandelt et al. 1999) implemented in Network 4.2.0.1; Fluxus Technology Ltd. For clarity, a cycle (not present in Figure 2.6) is completed with a red dashed line. Colored haplotypes are shared among sampling locations, white haplotypes are private, and black haplotypes are missing haplotypes. Haplotypes are labeled from 0–131, and haplotype sequences are listed in Appendix 2.

Figure 2.8: Multidimensional scaling based on mitochondrial DNA. Nei's genetic distances (d_{xy}) between sampling localities were calculated in Arlequin 3.1 (Excoffier et al. 2005). MDS was performed in R version 2.6.1 (Team 2007). Stress = 0.48 (calculated according to Venables and Ripley 1999 p. 333). The solid orange dots indicate North Island locations, the solid purple dots indicate South Island locations, the solid blue dot indicates the Chatham Islands, and the white dots outlined in purple indicate the two sampling locations from the north of the South Island (PHD and TCL, Figure 2.3).

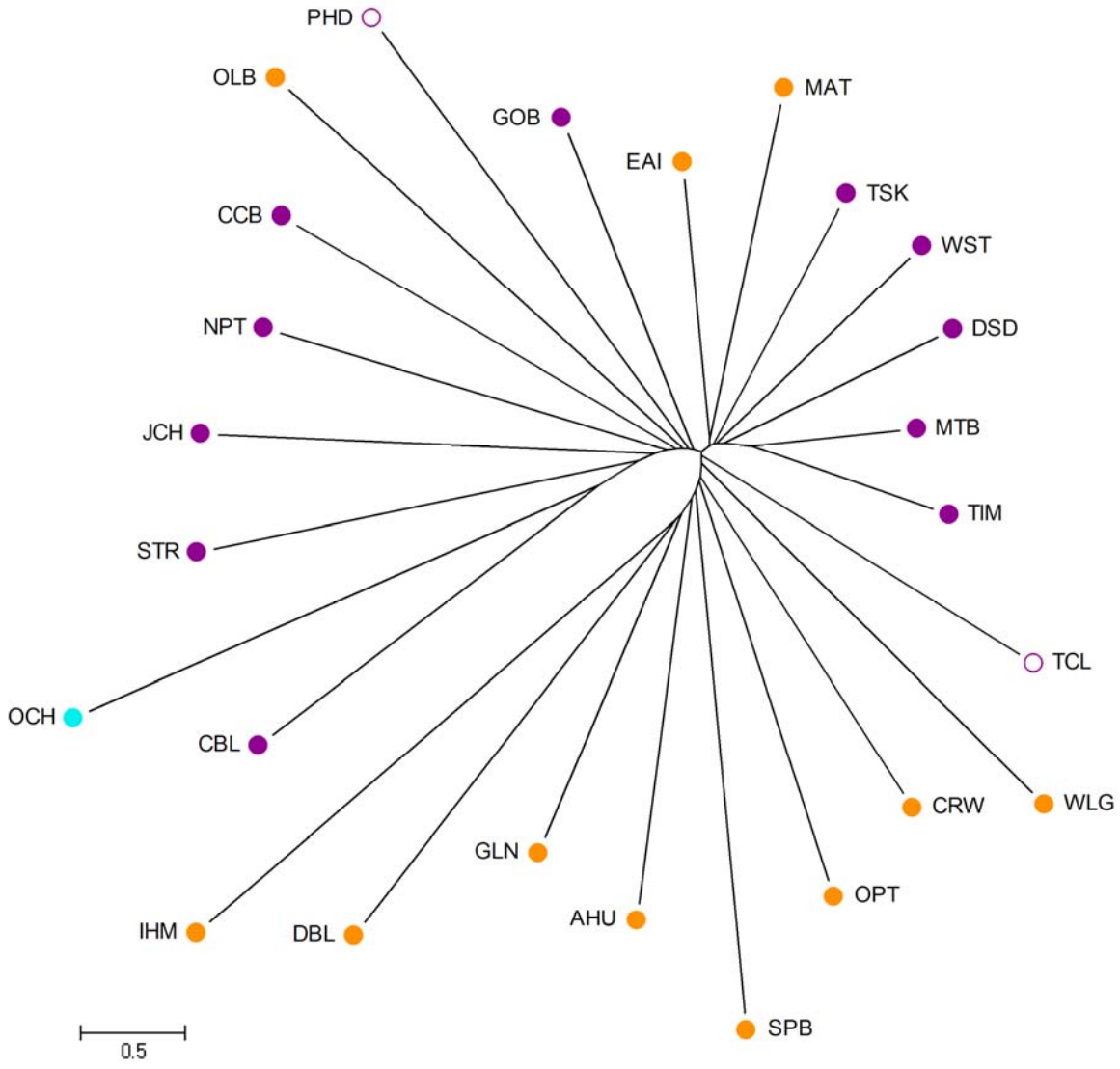


Figure 2.9: Neighbor joining analysis based on mitochondrial DNA. Nei's genetic distances (d_{xy}) between sampling localities were calculated in Arlequin 3.1 (Excoffier et al. 2005). The tree was built using the neighbor joining method (Saitou and Nei 1987) in MEGA4 (Tamura et al. 2007). The optimal tree with the sum of branch lengths = 48.266 is shown. The tree is drawn to scale. The solid orange dots indicate North Island locations, the solid purple dots indicate South Island locations, the solid blue dot indicates the Chatham Islands, and the white dots outlined in purple indicate the two sampling locations from the north of the South Island (PHD and TCL, Figure 2.3).

Table 2.5: Spatial analysis of molecular variance (SAMOVA) based on mitochondrial DNA. SAMOVAs were performed SAMOVA 1.0 (Dupanloup et al. 2002). K refers to the number of predefined groups used in the analyses. Statistical significances are indicated with asterisks (* $p < 0.05$, ** $p < 0.01$).

K	Φ_{ST}	Φ_{SC}	Φ_{CT}	Groups
2	0.15221**	0.03738**	0.11929*	[OCH] [AHU, CBL, CCB, CRW, DBL, DSD, EAI, GLN, GOB, IHM, JCH, MAT, MTB, NPT, OLB, OPT, PHD, SPB, STR, TCL, TIM, TSK, WLG, WST]
3	0.12445**	0.03052**	0.09689**	[OCH] [IHM] [AHU, CBL, CCB, CRW, DBL, DSD, EAI, GLN, GOB, JCH, MAT, MTB, NPT, OLB, OPT, PHD, SPB, STR, TCL, TIM, TSK, WLG, WST]
4	0.09965**	0.01661**	0.08444**	[OCH, CBL] [IHM] [DBL, GLN] [AHU, CCB, CRW, DSD, EAI, GOB, JCH, MAT, MTB, NPT, OLB, OPT, PHD, SPB, STR, TCL, TIM, TSK, WLG, WST]
5	0.08681**	0.00510**	0.08213**	[OCH, CBL] [IHM] [DBL, GLN] [MTB, TIM] [AHU, CCB, CRW, DSD, EAI, GOB, JCH, MAT, NPT, OLB, OPT, PHD, SPB, STR, TCL, TSK, WLG, WST]
6	0.08563**	0.00586**	0.08024**	[OCH, CBL] [IHM] [DBL] [GLN] [MTB, TIM] [AHU, CCB, CRW, DSD, EAI, GOB, JCH, MAT, NPT, OLB, OPT, PHD, SPB, STR, TCL, TSK, WLG, WST]
7	0.06603**	-0.01444**	0.07933**	[OCH, CBL] [IHM] [DBL, GLN] [MTB, TIM] [EAI] [AHU, CRW, OPT, SPB] [CCB, DSD, GOB, JCH, MAT, NPT, OLB, PHD, STR, TCL, TSK, WLG, WST]
8	0.06581**	-0.01461**	0.07926**	[OCH, CBL] [IHM] [DBL] [GLN] [MTB, TIM] [EAI] [AHU, CRW, OPT, SPB] [CCB, DSD, GOB, JCH, MAT, NPT, OLB, PHD, STR, TCL, TSK, WLG, WST]
9	0.06430**	-0.01475**	0.07789**	[OCH, CBL] [IHM] [DBL] [GLN] [MTB, TIM] [EAI] [AHU, SPB] [CRW, OPT] [CCB, DSD, GOB, JCH, MAT, NPT, OLB, PHD, STR, TCL, TSK, WLG, WST]
10	0.06381**	-0.01369**	0.07646**	[OCH] [CBL] [IHM] [DBL] [GLN] [MTB, TIM] [EAI] [AHU, SPB] [CRW, OPT] [CCB, DSD, GOB, JCH, MAT, NPT, OLB, PHD, STR, TCL, TSK, WLG, WST]
Excluding OCH				
2	0.10577**	0.03124**	0.07693**	[IHM] [AHU, CBL, CCB, CRW, DBL, DSD, EAI, GLN, GOB, JCH, MAT, MTB, NPT, OLB, OPT, PHD, SPB, STR, TCL, TIM, TSK, WLG, WST]
3	0.08860**	0.01904**	0.07091**	[IHM] [MTB, TIM] [AHU, CBL, CCB, CRW, DBL, DSD, EAI, GLN, GOB, JCH, MAT, NPT, OLB, OPT, PHD, SPB, STR, TCL, TSK, WLG, WST]
4	0.07986**	0.01018**	0.07040**	[IHM] [MTB, TIM] [DBL, GLN] [AHU, CBL, CCB, CRW, DSD, EAI, GOB, JCH, MAT, NPT, OLB, OPT, PHD, SPB, STR, TCL, TSK, WLG, WST]
5	0.07618**	0.00712**	0.06956**	[IHM] [MTB, TIM] [DBL, GLN] [CBL] [AHU, CCB, CRW, DSD, EAI, GOB, JCH, MAT, NPT, OLB, OPT, PHD, SPB, STR, TCL, TSK, WLG, WST]
6	0.06078**	-0.01094**	0.07094**	[IHM] [MTB, TIM] [DBL, GLN] [EAI] [AHU, CRW, OPT, SPB] [CBL, CCB, DSD, GOB, JCH, MAT, NPT, OLB, PHD, STR, TCL, TSK, WLG, WST]
7	0.06049**	-0.01094**	0.07066**	[IHM] [MTB, TIM] [DBL] [GLN] [EAI] [AHU, CRW, OPT, SPB] [CBL, CCB, DSD, GOB, JCH, MAT, NPT, OLB, PHD, STR, TCL, TSK, WLG, WST]
8	0.05721**	-0.01297**	0.06928**	[IHM] [MTB, TIM] [DBL, GLN] [EAI] [AHU, CRW, OPT] [SPB] [CBL] [CCB, DSD, GOB, JCH, MAT, NPT, OLB, PHD, STR, TCL, TSK, WLG, WST]
9	0.04666**	-0.02370**	0.06873**	[IHM] [DSD, MTB, TIM] [DBL, GLN] [EAI] [AHU, SPB] [CRW, OPT] [MAT, TSK, WST] [TCL, WLG] [CBL, CCB, GOB, JCH, NPT, OLB, PHD, STR]
10	0.04564**	-0.02593**	0.06976	[IHM] [MTB, TIM] [DBL, GLN] [EAI] [AHU, SPB] [CRW, OPT] [OLB] [PHD] [DSD, MAT, TCL, TSK, WLG, WST] [CBL, CCB, GOB, JCH, NPT, STR]

Table 2.6: AMOVA results based on mitochondrial DNA. AMOVAs tested the proposed groups listed in Table 2.2. AMOVAs were based on pairwise distances and implemented in Arelquin 3.1 (Excoffier et al. 2005). *p < 0.05, **p < 0.01.

Groups	Φ_{ST}	Φ_{SC}	Φ_{CT}
All samples	0.04453**		
Chatham Islands North and South Island	0.15221**	0.03738**	0.11929*
Chatham Islands North Island South Island	0.06407**	0.01976**	0.04521**
Chatham Islands North Island and north of South Island Remaining South Island	0.06375**	0.02091**	0.04376*
Excluding Chatham Islands			
North and South Island	0.03815**		
North Island South Island	0.05390**	0.02039*	0.03421**
North Island and north of South Island Remaining South Island	0.05336**	0.02155**	0.03251**

Table 2.7: Polymorphism data and neutrality results for mitochondrial DNA across AMOVA groupings. Locations within groups were specified in Table 2.2. Results generated in Arelequin 3.1 (Excoffier et al. 2005).

Group	N	Number of polymorphic sites	Number of haplotypes	Haplotype diversity $h \pm$ S.D.	Nucleotide diversity $\pi \pm$ S.D.	Tajima's D (p-value)	Fu's F_s (p value)
All locations	477	119	132	0.8990±0.0081	0.003827±0.002117	-2.25067 (0.000)	-25.16007 (0.000)
North & South Islands	464	116	129	0.8981±0.0084	0.003768±0.002089	-2.24785 (0.000)	-25.20583 (0.000)
North Island	222	97	97	0.9462±0.0085	0.004098±0.002252	-2.23390 (0.000)	-25.41254 (0.000)
South Island	242	51	45	0.8292±0.0131	0.003337±0.001887	-1.70454 (0.013)	-25.68839 (0.000)
North Island & north South Island	262	104	108	0.9434±0.0082	0.004110±0.002257	-2.24522 (0.001)	-25.31343 (0.000)
Remainder South Island	202	36	28	0.8071±0.0140	0.003178±0.001812	-1.30535 (0.074)	-9.05608 (0.015)
Chatham Islands	13	12	6	0.8205±0.0817	0.004890±0.002847	1.35874 (0.932)	1.36956 (0.764)

Table 2.8: Pairwise Φ_{ST} based mitochondrial DNA. Pairwise Φ_{ST} indices were calculated in Arelequin 3.1 (Excoffier et al. 2005). Significances were tested with 16002 permutations. Bold blue text indicates $p < 0.05$, and bold red text indicates values that are significant after standard Bonferroni corrections ($p < 0.00017$). $\Phi_{ST} = 0.000$ was used for $\Phi_{ST} < 0.0005$. Orange samples with black text are North Island locations. The blue sample with black text is the Chatham Islands. Purple samples with white text are South Island locations. White samples with purple text are from the north of the South Island (Figure 2.3).

	SPB	DBL	OPT	CRW	EAI	OLB	MAT	GLN	AHU	IHM	WLG	PHD	TCL	CBL	GOB	TSK	MTB	TIM	NPT	STR	CCB	DSD	JCH	WST
DBL	-0.013																							
OPT	-0.026	0.014																						
CRW	-0.019	0.012	-0.032																					
EAI	0.028	**0.115**	0.010	0.034																				
OLB	0.045	0.056	0.044	0.036	**0.065**																			
MAT	0.000	0.046	-0.006	-0.003	-0.009	0.004																		
GLN	0.019	-0.014	0.039	0.048	**0.139**	**0.073**	**0.057**																	
AHU	-0.025	-0.019	-0.022	-0.022	**0.057**	0.042	0.008	-0.006																
IHM	0.005	-0.005	0.028	0.037	**0.137**	**0.091**	**0.076**	0.017	0.003															
WLG	-0.005	0.010	-0.013	-0.019	**0.037**	0.005	-0.012	0.019	-0.019	0.033														
PHD	0.008	0.029	0.004	-0.002	0.043	-0.013	-0.003	0.050	0.005	0.042	-0.016													
TCL	0.011	0.025	0.004	-0.002	0.042	-0.005	-0.015	0.019	-0.006	**0.055**	-0.026	-0.015												
CBL	**0.093**	**0.089**	**0.107**	**0.107**	**0.162**	-0.010	0.078	**0.105**	**0.088**	**0.108**	0.049	0.001	0.023											
GOB	0.037	0.066	0.033	0.045	0.047	0.000	0.011	**0.066**	0.034	**0.084**	0.008	-0.014	-0.014	-0.008										
TSK	0.040	**0.104**	0.021	0.025	-0.005	0.008	-0.026	**0.116**	**0.048**	**0.125**	0.001	-0.001	-0.004	0.075	0.002									
MTB	**0.126**	**0.228**	**0.102**	**0.102**	**0.069**	**0.095**	**0.047**	**0.260**	**0.151**	**0.235**	**0.076**	**0.092**	**0.103**	**0.243**	**0.159**	0.018								
TIM	**0.118**	**0.203**	**0.093**	**0.082**	**0.075**	0.062	0.037	**0.231**	**0.135**	**0.219**	**0.057**	**0.066**	**0.071**	**0.185**	**0.126**	0.010	-0.018							
NPT	**0.056**	**0.075**	**0.058**	0.054	**0.074**	-0.025	0.016	**0.084**	0.054	**0.100**	0.012	-0.019	-0.009	-0.037	-0.028	0.002	**0.123**	**0.083**						
STR	0.041	0.043	0.046	0.041	**0.106**	-0.026	0.030	0.064	0.033	0.059	0.001	-0.035	-0.013	-0.056	-0.025	0.030	**0.173**	**0.122**	-0.045					
CCB	0.031	0.046	0.028	0.018	**0.084**	-0.018	0.017	0.062	0.021	0.053	-0.012	-0.037	-0.016	-0.022	-0.013	0.014	**0.124**	**0.085**	-0.031	-0.054				
DSD	**0.063**	**0.114**	0.044	0.028	0.062	-0.009	0.005	**0.142**	0.064	**0.139**	0.001	-0.007	-0.011	0.046	0.021	-0.018	0.061	0.000	-0.009	0.004	-0.008			
JCH	0.046	0.056	0.050	0.050	**0.087**	-0.019	0.024	0.062	0.039	**0.068**	0.007	-0.025	-0.006	-0.041	-0.030	0.018	**0.149**	**0.114**	-0.041	-0.053	-0.039	0.018		
WST	0.036	**0.088**	0.020	0.021	0.011	-0.016	-0.026	**0.101**	0.040	**0.108**	-0.013	-0.021	-0.023	0.032	-0.022	-0.052	0.037	0.013	-0.029	-0.006	-0.016	-0.038	-0.012	
OCH	**0.157**	**0.140**	**0.188**	**0.194**	**0.258**	0.043	**0.162**	**0.163**	**0.160**	**0.148**	**0.123**	0.061	**0.106**	-0.046	0.063	**0.170**	**0.346**	**0.291**	0.019	-0.010	0.037	**0.153**	0.001	0.122

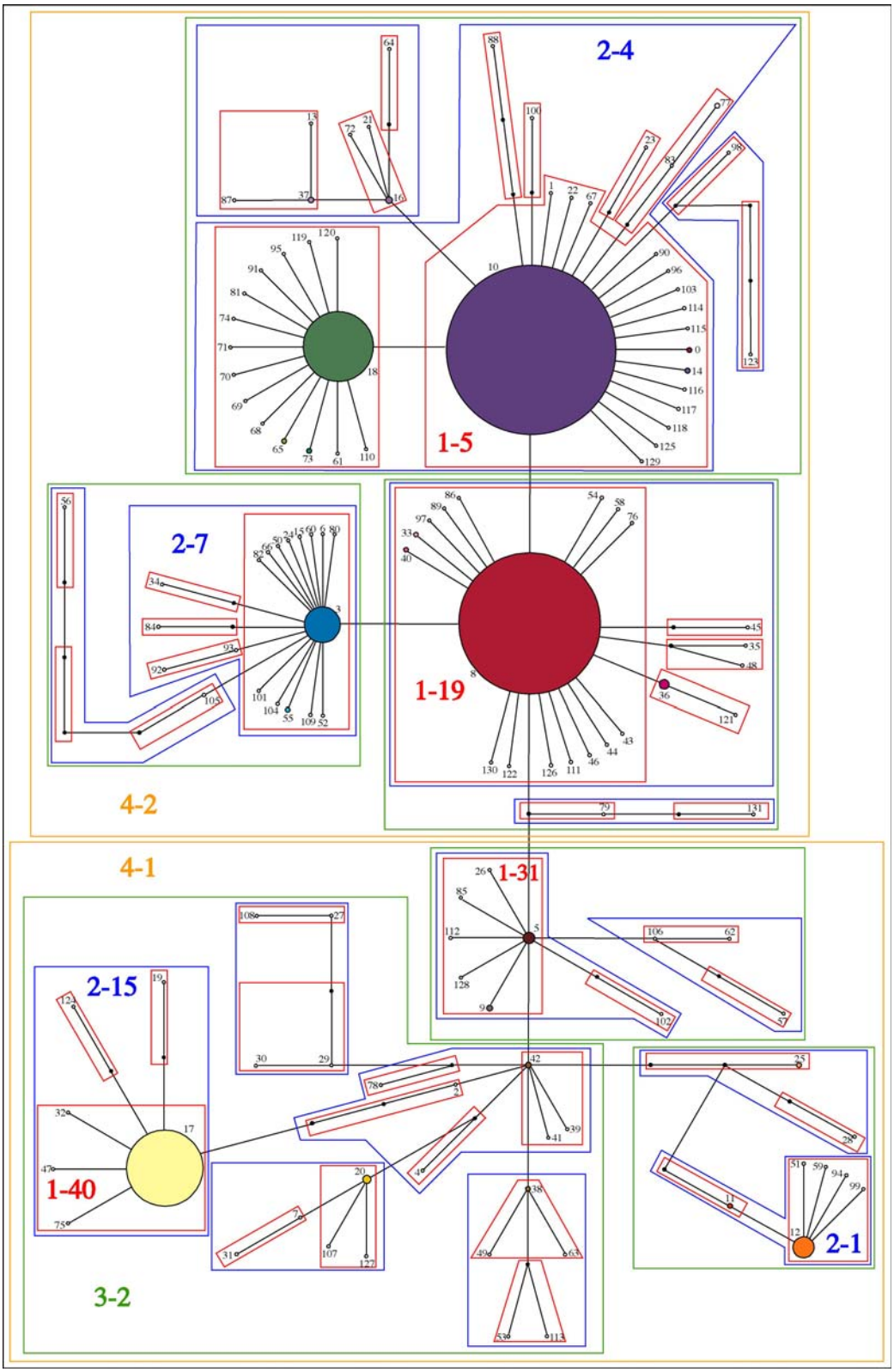


Figure 2.10: Nested network for Nested Clade Phylogeographic Analysis (NCPA). Original statistical parsimony network was presented in Figure 2.6. 1-step clades are red, 2-step clades are blue, 3-step clades are green, and 4-step clades are orange. The entire network was nested in 5 steps (black). Only clades listed in Table 2.7 are numbered. All other clades were inferred to be panmictic.

Table 2.9: Interpretation of Nested Clade Phylogeographic Analysis. Inferences used the November 11, 2005 inference key supplied with GeoDis 2.5 (Posada et al. 2006).

Clade	Chain of inference	Demographic event inferred
1-5	1-2-11-17-4-No	Restricted gene flow with isolation by distance
1-19	1-2-11-17-4-No	Restricted gene flow with isolation by distance
1-31	1-2-11-17-No	Inconclusive outcome
1-40	1-2-11-17-No	Inconclusive outcome
2-1	1-2-11-17-No	Inconclusive outcome
2-4	1-2-3-5-6*-7-8-No	Sampling design inadequate to discriminate between isolation by distance (short distance movements) versus long distance dispersal *Insufficient genetic resolution to discriminate between range expansion/colonization and restricted dispersal/gene flow
2-7	1-2-3-4-No	Restricted gene flow with isolation by distance
2-15	1-2-3-4-No	Restricted gene flow with isolation by distance
3-2	1-2-11-12-13-14-Yes	Sampling design inadequate to discriminate between contiguous range expansion, long distance colonization, and past fragmentation
4-1	1-2-3-5-6-Too few clades	Insufficient genetic resolution to discriminate between range expansion/colonization and restricted dispersal/gene flow
4-2	1-2-3-4-No	Restricted gene flow with isolation by distance

Table 2.10: Mantel tests using mitochondrial DNA. Mantel tests were performed within each group proposed in Table 2.2 and implemented in Arlequin 3.1 (Excoffier et al. 2005).

Group	Correlation coefficient (p-value)
All locations	0.189 (0.047)
North & South Islands	0.119 (0.126)
North Island	0.229 (0.054)
South Island	-0.162 (0.966)
North Island & north South Island	0.237 (0.047)
Remainder South Island	-0.132 (0.825)
Chatham Islands	NA

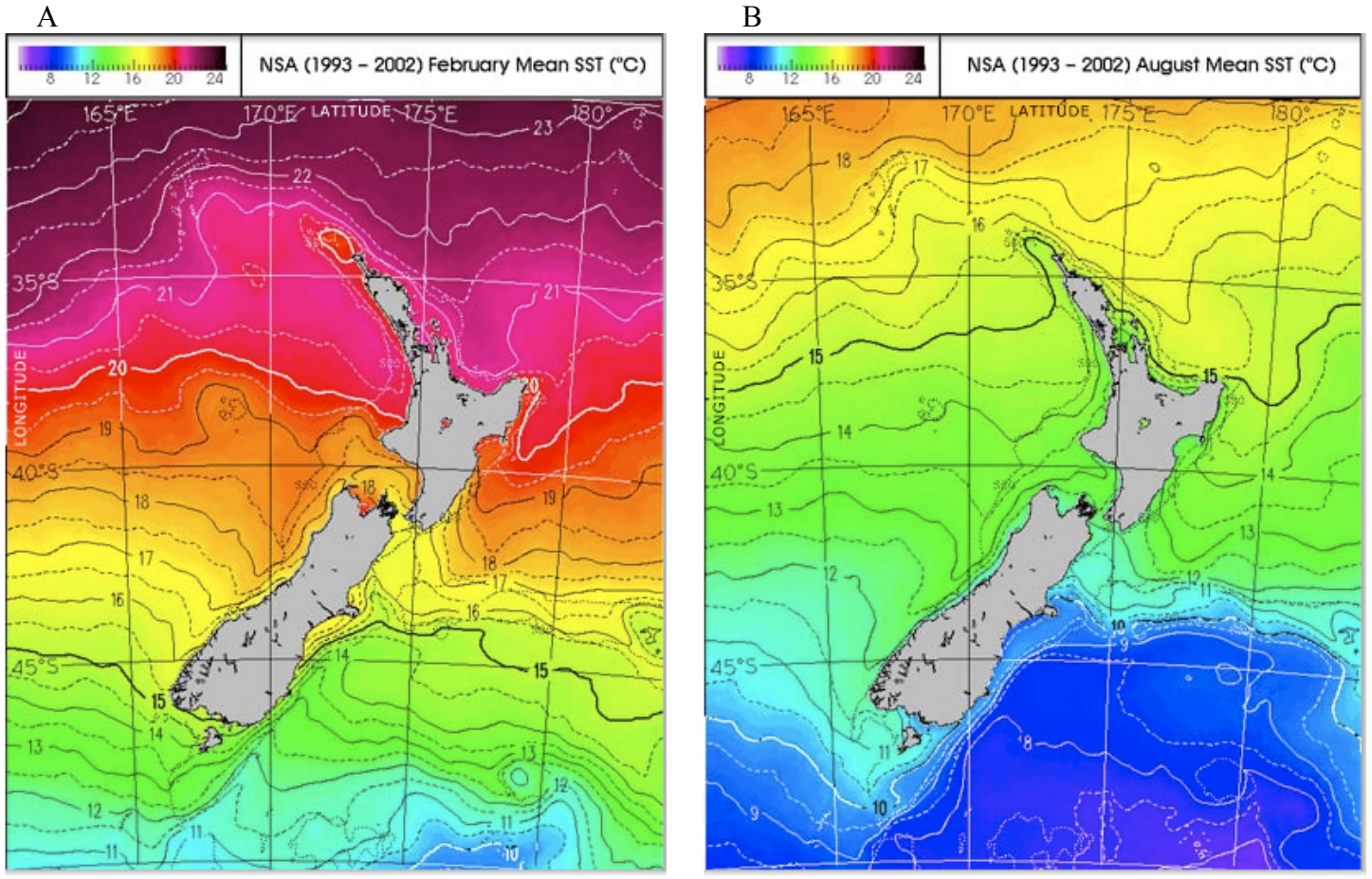


Figure 2.11: Mean sea surface temperatures around New Zealand from 1993–2002. A) Mean sea surface temperatures (°C) in summer (February). B) Mean sea surface temperatures (°C) in winter (August). Images were produced by the National Institute of Water and Atmosphere (NIWA) for The Encyclopedia of New Zealand and can be accessed at www.teara.govt.nz.

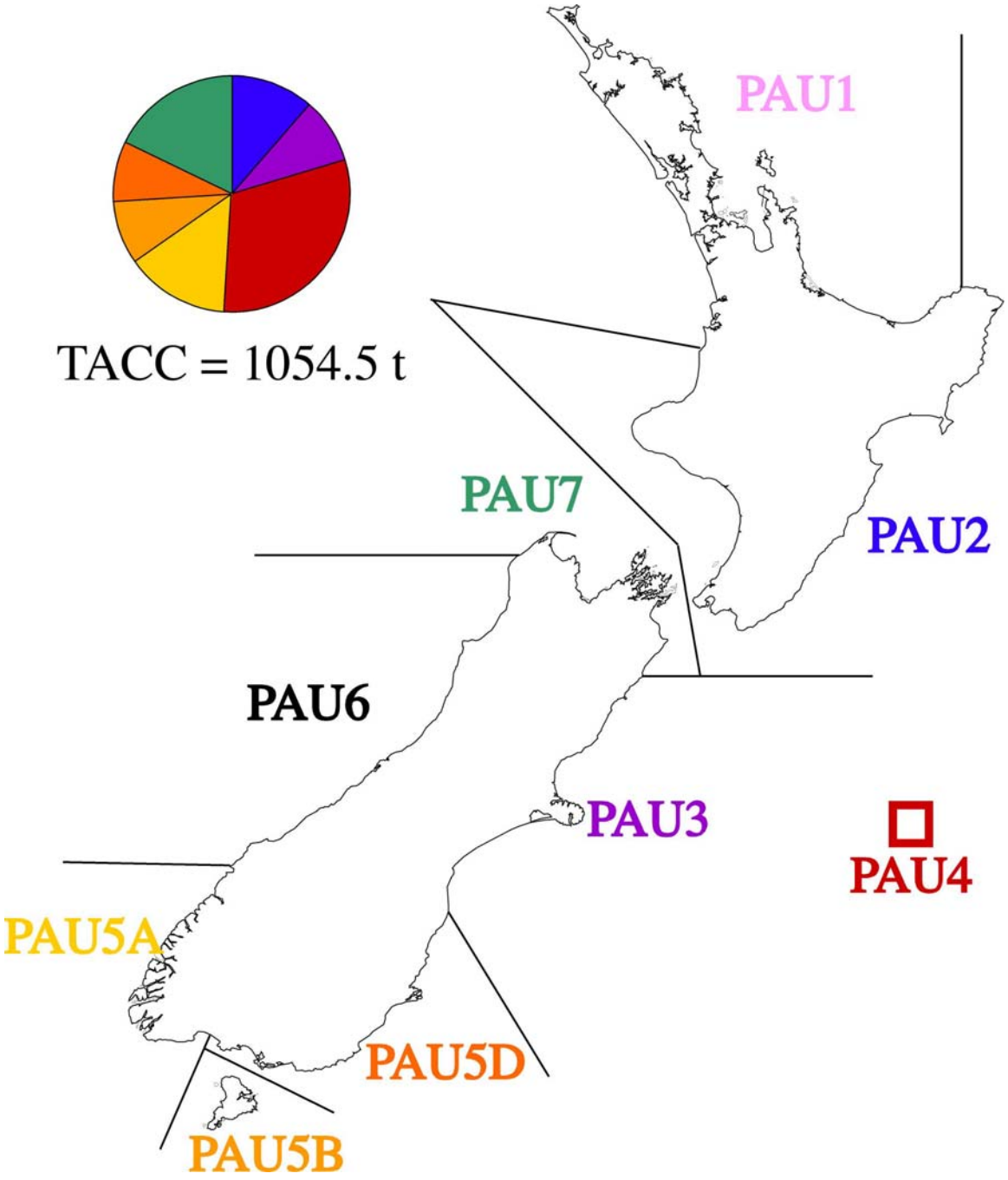
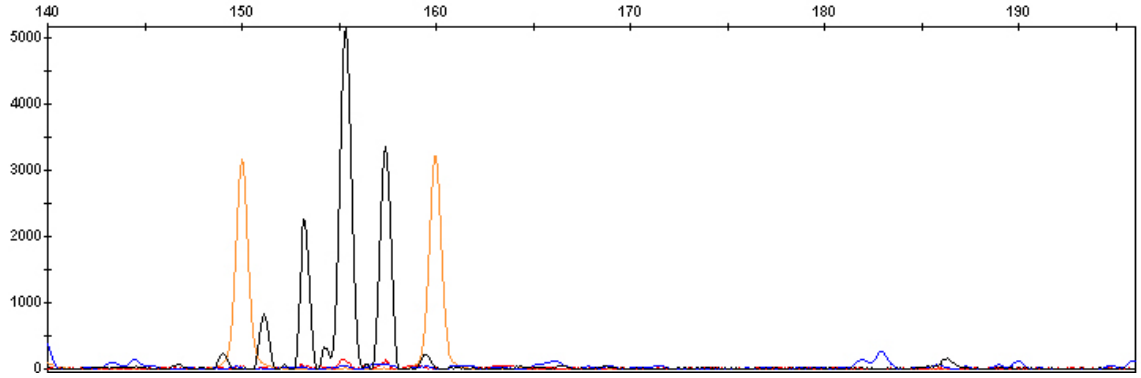


Figure 2.12: *Haliotis iris* fishing management areas and Total Allowable Commercial Catch (TACC) per area for the 2008–2009 fishing season. PAU6 and PAU10 (Kermadecs, not shown) each have 1 t TACC. However, no reported landings occurred from 2000–2006 for PAU6 or ever for PAU10. For the 2006–2007 fishing year, 100% of the TACC was reported caught for PAU6, and only 0.76 t of the 1.93 t TACC were captured. Landings for all other areas reported captures greater than 94% of the TACC for 2006–2007. Therefore, PAU1, PAU6, and PAU10 are not included in the pie diagram. Data from www.fish.govt.nz.

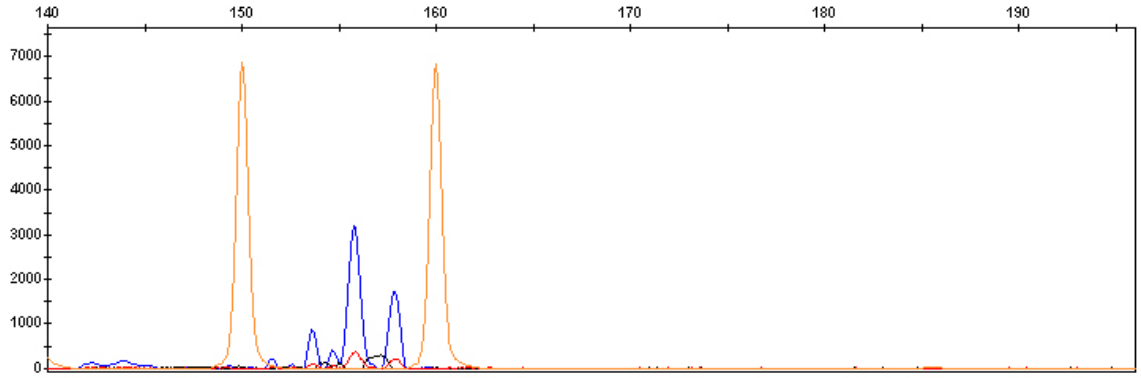
Table 3.1: Loci used for testing cross-species amplification on *H. iris* DNA. Listed are the original species for which a locus was developed, the locus name according to the original reference, the reference for the primer sequences, the expected allele size range for the original species, results of amplification in *H. iris* (band, one or two bands present; lots of bands, three or more bands present; messy, many bands or smears present), screening method (Agarose, 2% agarose gels stained with ethidium bromide; AB3100, genotyped with fluorescently labeled dUTPs), annealing temperature and $MgCl_2$ concentrations for primers with clean amplification, and whether an optimized locus produced a usable microsatellite.

Original species	Locus	Reference	Expected size range	Amplfication	Screening method	Clean amplification	Marker suitability
H. rubra	CmrHr 2.20	(Evans et al. 2001)	186	Messy	Agarose	–	–
H. rubra	CmrHr 2.23	(Evans et al. 2001)	258-266	Messy	Agarose, AB3100	–	–
H. rubra	CmrHr 2.27	(Evans et al. 2001)	346	Messy	Agarose	–	–
H. rubra	CmrHr 2.30	(Evans et al. 2001)	284-328	Messy	Agarose	–	–
H. rubra	CmrHr 2.36	(Evans et al. 2001)	83-121	Messy	Agarose, AB3100	–	–
H. rubra	CmrHr 2.3	(Evans et al. 2001)	100	Messy	Agarose, AB3100	–	–
H. rubra	CmrHr 2.22	(Evans et al. 2001)	117-193	Messy	Agarose, AB3100	–	–
H. rubra	CmrHr 1.24	(Evans et al. 2001)	216-236	Messy	Agarose, AB3100	–	–
H. midae	HmD11	(Bester et al. 2004)	292-352	Messy	Agarose, AB3100	–	–
H. midae	HmD14	(Bester et al. 2004)	142-180	Messy	Agarose	–	–
H. midae	HmD30	(Bester et al. 2004)	124-150	Messy	Agarose, AB3100	–	–
H. midae	HmD36	(Bester et al. 2004)	220-304	Messy	Agarose	–	–
H. midae	HmD55	(Bester et al. 2004)	183-211	Messy	Agarose, AB3100	–	–
H. midae	HmD59	(Bester et al. 2004)	106-150	Band	Agarose, AB3100	55 °C 2.5 mM $MgCl_2$	Monomorphic peak at position 220; $(AG)_N$
H. midae	HmD60	(Bester et al. 2004)	155-171	Messy	Agarose, AB3100	–	–
H. midae	HmD61	(Bester et al. 2004)	234-298	Messy	Agarose, AB3100	–	–
H. midae	HmD33	(Bester et al. 2004)	129-205	Messy	Agarose, AB3100	–	–
H. midae	HmSP1	(Bester et al. 2004)	192-276	Messy	Agarose, AB3100	–	
H. midae	HmSP5	(Bester et al. 2004)	185-219	Band	Agarose, AB3100	55 °C 2.5 mM $MgCl_2$	Monomorphic peak at position 412; No repeat present in sequence

A



B



C

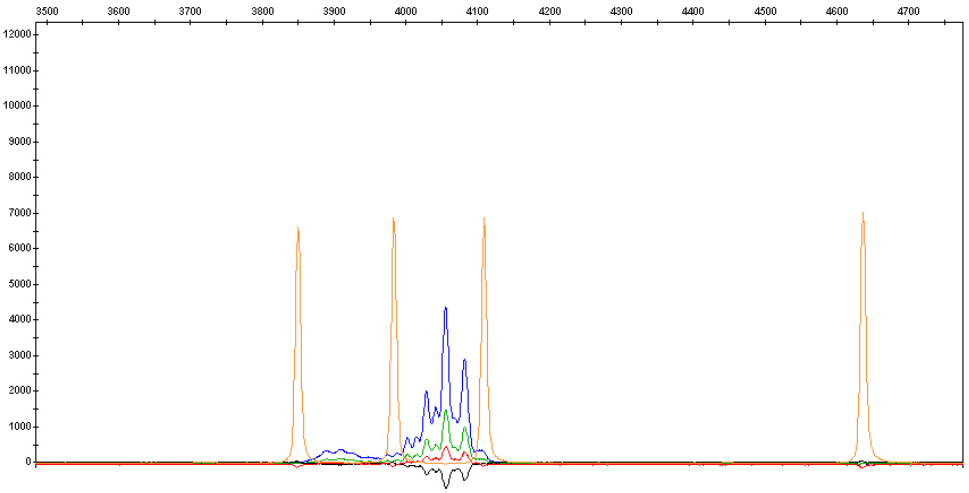


Figure 3.1: Example electropherogram for a microsatellite labeled with dUTPs. Amplification was of European hedgehog (*Erinaceus eupaeus*) locus EEU6 (Becher and Griffiths 1997). Microsatellite amplification of the same allele with A) a fluorescently labeled primer and B) fluorescently labeled dUTPs. C) The raw data image of B showing a spiky shape. Images were produced with GeneMarker v1.6 Demo (SoftGenetics®). Data was kindly provided by Dr. Marie Hale.

Table 3.2: Primer pairs for polymorphic microsatellite loci isolated by ATG genetics, Inc.. Primers named R.2 (e.g., AB10R.2) are primers that were redesigned using Primer 3 (Rozen and Skaletsky 2000). Listed are the loci names, primer sequences, methods used for visualizing and optimizing PCRs, and primers ordered with fluorescent tags. Colors for fluorescent tags were chosen based on allele sizes of each locus and with the intent of forming two pooling groups to run on an ABI3100. Applied Biosystems Pty. Ltd. (Foster City, CA USA) supplied primers labeled with the VIC, NED and PET tags. Sigma-Aldrich Pty. Ltd. (Castle Hill, NSW Australia) supplied primers labeled with the 6-FAM tag. Loci sequences are listed in Appendix 6.

Locus	Primer (5′-3′)	Agarose	dUTPs	Fluorescent tagged M13 primers	Fluorescent tagged primers
AB1R	F-ATGTGTGCTTGCGCGTGAGTG	X	X	X	NED
	R-GACAACTGGACGCGTCCATAATG	X	X	X	
AB3R	F-CACACACGCCCGAACGCACAG	X	X		PET
	R-CGTCACACTGACACAGCGAG	X	X		
AB5R	F-TTAGCGCTTTGAGCATACACC	X	X	X	PET
	R-CTGCAACACCACCTTCGTCC	X	X	X	
AB10R	F-ATTGCTGGCAGAATATCAATAG	X	X		Not ordered
	R-TCAATCTCTCTGACTTGGCAG	X	X		
AB10R.2	F-TTTCAGTCCCTGGTTTCCTG	X	X		
	R-TGCTGGCAGAATATCAATAGAA	X	X		
AB11R	F-ACCGTAATTTTGAACATTATC	X	X		VIC
(= AB25R)	R-CGGAAGTAGCTGAAACGAATC	X	X		
AB13R	F-GAAAAATCGAGTTTGGCGATCC	X	X	X	
	R-GGTAGGAAATGGTTACATACTAAC	X	X	X	
AB13R.2	F-AACAGCAAGGTTGCGGTAAG	X			6-FAM
	R-AAATCGAGTTTGGCGATCC	X			
AB14R	F-TAAGTGAAGAGACTGACGCTTG	X	X	X	VIC
	R-CATTAAGTTGAGGTTCAGGGCTG	X	X	X	
AB17R	F-TTCTGCTATGGATCTGATACC	X	X		
	R-AATGGGACGGTAGGGTAGCC	X	X		
AB17R.2	F-TTACGGCTAAGATCCGGGTA	X	X		6-FAM
	R-TGCTTTTCCTTCTGCTATGGA	X	X		
AB21R	F-GTGGTTTCATTGGTTATGTTTC	X	X	X	
(= AB22R)	R-GTCATGTATATACGTCGGTTC	X	X	X	
AB21R.2	F-ATATACGTCGGTTCCGTGCT	X	X		NED
	R-AAGGGGTGGTTTCATTGGTT	X	X		
AB22R	F-GTGGTTTCATTGGTTATGTTTC	X	X		Not ordered
(= AB21R)	R-CTACTACTGGCCTTCTAGGTC	X	X		
AB23R	F-GTAATCTTCGGTAATTACTGG	X	X	X	
	R-CCAATCTCATATCAGCGTCAG	X	X	X	
AB23R.2	F-GCCAATCTCATATCAGCGTCA	X			VIC
	R-TTCGGTAATTACTACTGGAGGAAA	X			
AB24R	F-CCACCAAGGATGGCAATGCC	X	X		Not ordered
	R-GTAATCTTCGGTAATTACTGG	X	X		
AB24R.2	F-GATGGCAATGCCTCAGCTAC	X	X		
	R-CCTTCTACCCAGCTTCTATCTG	X	X		
AB25R	F-CAAGTGACGCAGTTATCAAA	X	X		Not ordered
(= AB11R)	R-CCGTAATTTTGAACATTATCTTA	X	X		
AB30R	F-TCGGCGAAGACTGTACTGCC	X	X	X	
	R1-AAACCCTCTTGATATCTGTCC	X	X	X	
	R2-TGGGGTGATGATTGTGATGC	X	X	X	
AB30R.2	F-CTGGGGTGATGATTGTGATG	X	X		6-FAM
	R-AATCACTCGGCGAAGACTGT	X	X		
AB31R	F-CATTGTTTTGGTATGCGATATGG	X	X		PET
	R-CTCTCATAACTGTTTTAGAGTGG	X	X		

Table 3.3: Preliminary screening of 13 *H. iris* microsatellite loci. Listed are the locus names (primers sequences given in Table 3.2), optimum annealing temperatures (T_A, T_a; PCR conditions were similar to those given in Table 3.4), number of individuals screened (N), number of alleles identified (A), observed size range, and observed heterozygosity (H_O). Suitability refers to whether markers could be pursued given a limited time frame. Duplicate loci meant that more than two alleles were amplified in a single individual (Figure 3.3A). Large allele dropout referred to the lack of amplification of large alleles due to the preferential amplification of shorter alleles (Figure 3.3B). Alleles larger than 500 bp could not be accurately measured with GeneScan™-500 LIZ® Size Standard. Allele shape meant that alleles were difficult to score because of stutter patterns and different shapes (Figure 3.3C).

Locus	T_A, T_a (°C)	N	A	Size range (bp)	H_O	Suitability
AB1	65, 65	18	23	162.9–350.9	1.00	Duplicate loci, different shaped alleles
AB3	60, 60	7	8	215.2–287.0	1.00	Duplicate loci
AB5	57, 55	6	10	199.6–344.2	1.00	Large allele dropout
AB11	57, 57	7	9	165.0–190.6	0.86	Different shaped alleles
AB13	60, 60	7	12	134.7–241.5	1.00	Large allele dropout
AB14	57, 57	7	9	173.0–212.4	1.00	Suitable
AB17	60, 60	7	11	191.9–539.4	0.86	Alleles larger than 500 bp
AB21	63, 63	7	11	179.3–259.7	1.00	Suitable
AB23	55, 55	6	11	179.2–262.3	0.83	Allele shapes difficult to score
AB30	60, 60	7	10	107.5–262.1	1.00	Large allele dropout
AB31	57, 57	7	10	174.3–205.8	1.00	Suitable

Table 3.4: Optimized PCR conditions for loci AB14, AB21, and AB31. T_A and T_a refer to annealing temperatures in Table 3.3. Primer sequences were given in Table 3.2. * indicates fluorescent tagged primer.

Locus	AB14	AB21	AB31
Reagents			
Buffer	1X	1X	1X
$MgCl_2$	1.5 mM	1.5 mM	1.5 mM
dNTPs	200 µM	200 µM	200 µM
*F-primer	0.33 µM	0.33 µM	0.33 µM
R-primer	0.33 µM	0.33 µM	0.33 µM
Taq	0.6 units	0.6 units	0.6 units
Thermocycling profile			
12 minutes	95 °C	95 °C	95 °C
cycles	10	10	10
15 s	94 °C	94 °C	94 °C
15 s (T_A)	57 °C	63 °C	57 °C
15 s	72 °C	72 °C	72 °C
cycles	22	21	23
15 s	89 °C	89 °C	89 °C
15 s (T_a)	57 °C	63 °C	57 °C
15 s	72 °C	72 °C	72 °C
20 minutes	72 °C	72 °C	72 °C

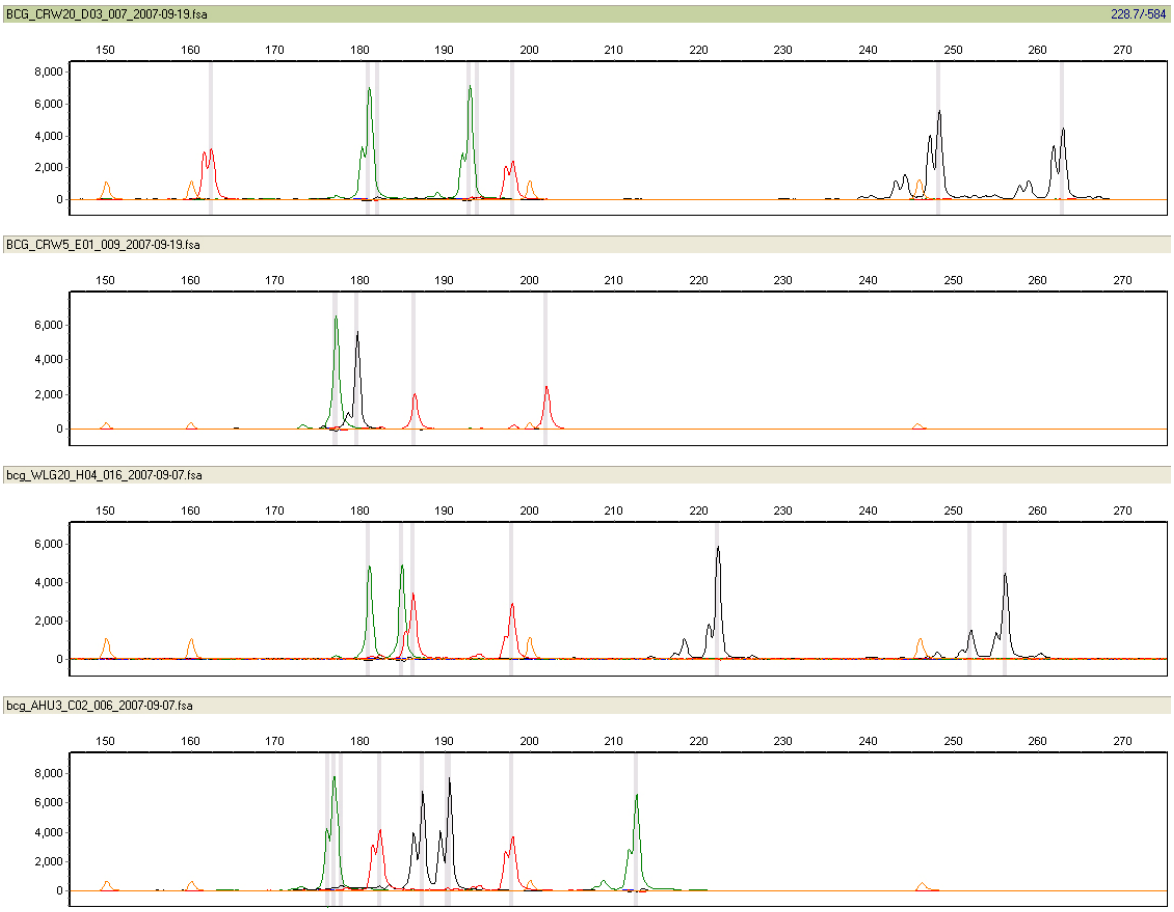
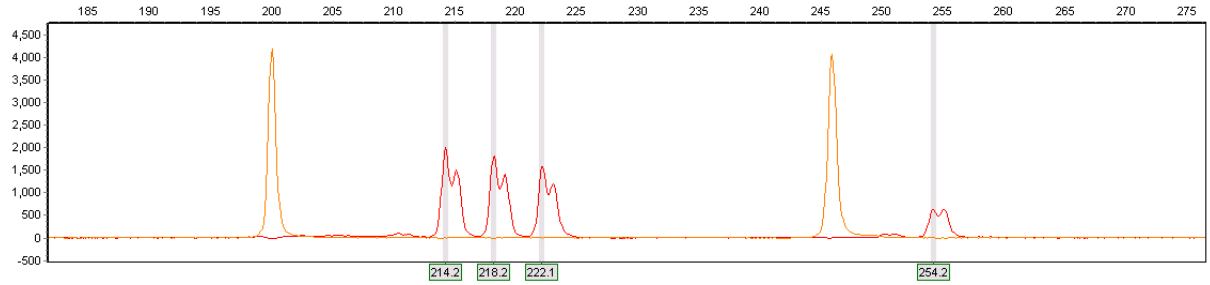
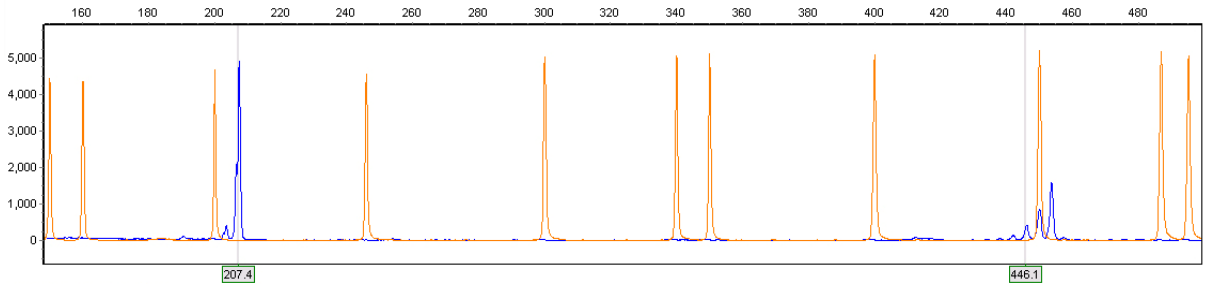


Figure 3.2: Examples of electropherograms for loci AB14, AB21, and AB31. The green peaks are locus AB14, the black peaks are locus AB21, and the red peaks are locus AB31.Orange peaks are GeneScan™-500 LIZ® Size Standard. Individual CRW5 (second electropherogram from the top) was homozygous for loci AB14 and AB21 and heterozygous for locus AB31. All other individuals pictured were heterozygous for all loci. Images were produced with GeneMarker v1.6 Demo (SoftGenetics®).

A



B



C

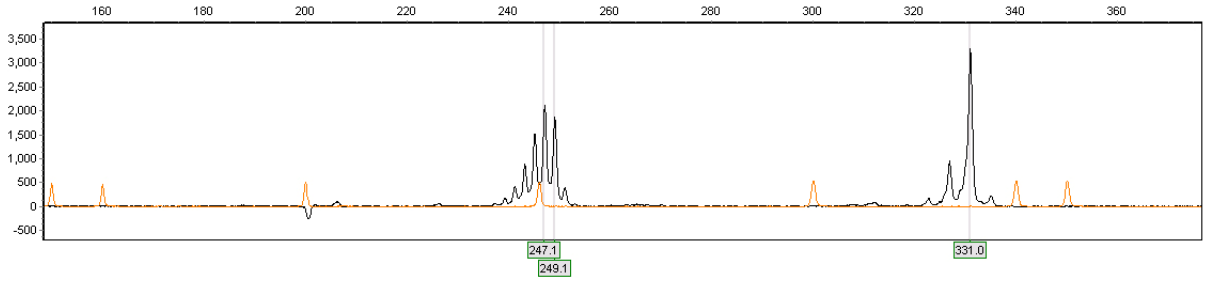


Figure 3.3: Examples of electropherograms for loci deemed unsuitable for further screening. Loci were determined inadequate for a larger scale population genetic study because (A) more than 2 alleles per individual (e.g., locus AB3), (B) large allele range (e.g., locus AB17), and (C) different shaped alleles (e.g., locus AB1). Images were produced with GeneMarker v1.6 Demo (SoftGenetics®).

Table 3.5: Genotyping error rates. Genotyping error rates were calculated according to Hoffman and Amos (2005) and Pompanon et al. (2005).

Locus	Number of successful reactions	Number of mistyped reactions	Number of mis-typed alleles	Error rate per reaction (%)	Error rate per allele (%)
AB14R	84	3	6	3.6	3.6
AB21R	74	3	6	4.1	4.1
AB31R	87	4	6	4.6	3.4
Overall	245	10	18	4.1	3.7

Table 3.6: Characterization of loci AB14, AB21, and AB31. Listed are the locus name (see Table 3.2 for primers), number of individuals that had successful amplifications, number of alleles identified (A), observed size range, observed heterozygosity (H_O), unbiased expected heterozygosity (H_E). Observed and unbiased expected heterozygosities were calculated in GeneAlEx 6.1 (Peakall and Smouse 2006). Null alleles were estimated according to Brookfield (1996) in MICRO-CHECKER 2.2.3 (Van Oosterhout et al. 2004). * indicates significant deviation from Hardy–Weinberg equilibrium, assessed using exact tests in GENEPOP 4.0 (Rousset 2008).

Locus	Repeat	N	A	Size range (bp)	H_O	H_E	Null allele frequency
AB14	$(CTAT)_N$	455	23	168.7–224.4	0.754	0.876*	0.0644
AB21	$(CTAT)_N$	447	84	166.9–287.1	0.895	0.973*	0.0379
AB31	$(CTAT)_N$	459	42	162–177.5	0.826	0.825*	0.0153

Table 4.1: Standard polymorphism indices for 25 sampling locations and three microsatellite loci. Listed are N, number of individuals; A, number of alleles; Ae, effective number of alleles; Ar, allelic richness; H_O, observed heterozygosity; and H_E, unbiased expected heterozygosity. All values except allelic richness were calculated in GeneAlEx 6.1 (Peakall and Smouse 2006). Allelic richness was calculated in Fstat 2.9.3 (Goudet 2002). Deviations from Hardy–Weinberg equilibrium were calculated in GENEPOP 4.0 (Rousset 2008). * indicates samples that were out of Hardy–Weinberg equilibrium for a single locus. ^{S} indicates sites that are out of Hardy–Weinberg equilibrium over all loci.

	AB14						**AB21**						**AB31**					
Site	**N**	**A**	**Ae**	**Ar**	**H_O**	**H_E**	**N**	**A**	**Ae**	**Ar**	**H_O**	**H_E**	**N**	**A**	**Ae**	**Ar**	**H_O**	**H_E**
AHU	22	12	7.683	9.724	0.818	0.890	22	26	20.167	18.811	1.000	0.973	22	21	13.444	15.637	0.909	0.947
CBL^S	14	8	5.521	7.921	0.714	0.849	14	19	14.519	18.061	0.857	0.966*	14	8	4.261	7.714	0.857	0.794
CCB	19	9	7.149	8.591	0.684	0.883*	17	22	17.515	18.581	1.000	0.971	19	8	5.597	7.800	0.895	0.844
CRW^S	19	9	6.505	8.024	0.895	0.869	19	28	22.563	20.915	0.789	0.982*	18	11	8.308	10.053	0.944	0.905
DBL	15	7	5.844	6.971	0.933	0.857	15	21	15.000	18.825	0.933	0.966	14	13	8.909	12.563	0.857	0.921
DSD	14	8	5.444	7.786	0.571	0.847*	14	18	13.067	17.132	1.000	0.958	14	11	6.533	10.571	1.000	0.878*
EAI	22	9	5.319	7.825	0.818	0.831	18	20	9.000	15.863	0.722	0.914*	23	11	6.451	9.497	0.913	0.864
GLN	20	10	6.400	8.679	0.750	0.865	20	28	22.857	20.639	0.950	0.981	20	15	8.247	11.846	0.900	0.901
GOB^S	19	10	8.699	9.564	0.684	0.909*	19	19	14.440	15.824	0.947	0.956	19	12	6.119	10.179	0.895	0.859
IHM^S	20	11	7.273	9.880	0.800	0.885	20	25	19.048	19.010	0.900	0.972	20	13	7.273	10.642	0.950	0.885
JCH^S	19	13	7.934	10.969	0.632	0.898*	20	21	14.286	16.526	0.800	0.954*	20	13	6.504	11.100	0.750	0.868
MAT	18	7	5.143	6.646	0.778	0.829	18	26	20.250	20.345	0.889	0.978	14	8	4.900	7.852	0.714	0.825
MTB^S	21	10	7.230	9.000	0.619	0.883*	22	24	17.286	17.535	0.955	0.964	22	10	5.348	8.415	0.773	0.832
NPT^S	19	8	6.278	7.622	0.842	0.863	20	22	18.605	17.732	0.900	0.971	19	8	3.297	6.931	0.632	0.716
OCH	13	5	3.634	5.000	0.615	0.754	13	22	18.778	22.000	1.000	0.985	13	12	9.389	12.000	0.923	0.929
OLB^S	17	9	7.918	8.877	0.765	0.900	19	20	15.042	16.195	0.895	0.959	19	9	4.599	7.942	0.737	0.804
OPT^S	19	9	5.918	8.153	0.737	0.853	18	22	16.200	17.785	0.778	0.965*	19	14	7.681	11.764	0.789	0.893
PHD^S	17	8	5.453	7.429	0.765	0.841	16	23	18.963	19.841	0.813	0.978*	17	8	3.613	7.241	0.647	0.745
SPB	21	10	7.113	8.666	0.905	0.880	20	25	19.048	18.775	0.900	0.972	21	15	8.092	12.081	0.905	0.898
STR	14	11	8.000	10.709	0.643	0.907*	13	17	11.655	17.000	1.000	0.951	15	9	4.128	8.558	0.800	0.784
TCL	19	8	5.967	7.599	0.789	0.855	20	22	17.021	17.218	1.000	0.965	20	12	6.838	9.898	0.800	0.876
TIM^S	22	11	7.170	9.515	0.864	0.881*	22	19	14.667	15.254	0.773	0.953*	23	9	4.600	7.717	0.783	0.800
TSK^S	18	10	6.680	9.006	0.667	0.875	13	18	15.364	18.000	0.846	0.972	19	12	3.722	9.373	0.789	0.751
WLG^S	19	10	5.918	8.413	0.789	0.853	20	25	17.391	18.779	0.950	0.967	20	10	5.333	8.440	0.650	0.833*
WST^S	15	8	6.522	7.732	0.600	0.876*	15	13	11.842	12.658	0.800	0.947*	15	10	5.422	9.319	0.867	0.844

Table 4.2: Pairwise F_{ST} based on microsatellites. Pairwise F_{ST} indices were calculated in Arlequin 3.1 (Excoffier et al. 2005). Significances were tested with 16002 permutations. Bold blue text indicates $p < 0.05$, and bold red text indicates values that are significant after standard Bonferroni corrections ($p < 0.00017$). $F_{ST} = 0.000$ was used for $F_{ST} < 0.0005$. Orange samples with black text are North Island locations. The blue sample with black text is the Chatham Islands. Purple samples with white text are South Island locations. White samples with purple text are from the north of the South Island (Figure 2.3). Orange samples with black text are North Island locations. The blue sample with black text is the Chatham Islands. Purple samples with white text are South Island locations. White samples with purple text are from the north of the South Island (Figure 2.3).

	SPB	DBL	OPT	CRW	EAI	OLB	MAT	GLN	AHU	IHM	WLG	PHD	TCL	CBL	GOB	TSK	MTB	TIM	NPT	STR	CCB	DSD	JCH	WST
DBL	0.000																							
OPT	-0.006	-0.002																						
CRW	-0.004	0.003	0.006																					
EAI	-0.001	-0.009	-0.011	-0.001																				
OLB	0.001	0.012	-0.002	0.011	0.011																			
MAT	-0.017	0.004	-0.020	0.004	-0.017	-0.011																		
GLN	-0.009	0.000	-0.007	0.002	0.002	0.005	-0.011																	
AHU	0.001	0.000	0.000	0.002	0.001	**0.019**	-0.004	0.000																
IHM	-0.004	0.003	-0.001	0.003	0.004	0.002	0.000	-0.005	0.005															
WLG	0.000	-0.001	0.003	0.005	-0.008	0.001	-0.012	0.009	0.011	0.009														
PHD	0.003	0.020	-0.006	0.017	0.012	-0.014	-0.017	0.003	**0.025**	0.006	0.006													
TCL	0.005	0.001	0.001	0.009	-0.007	**0.017**	0.002	0.013	0.008	0.004	0.005	**0.021**												
CBL	**0.015**	**0.023**	0.003	**0.023**	0.009	-0.010	-0.010	**0.022**	**0.022**	**0.017**	0.005	0.000	**0.018**											
GOB	0.006	0.013	0.003	0.004	0.009	-0.011	-0.001	0.012	**0.017**	0.009	0.008	0.002	0.013	0.003										
TSK	-0.007	**0.021**	0.003	-0.007	**0.031**	-0.020	-0.012	0.001	**0.014**	-0.003	-0.005	-0.005	0.016	-0.005	-0.015									
MTB	0.000	0.012	0.003	0.006	0.005	-0.003	-0.009	0.007	**0.016**	0.012	0.004	0.001	**0.014**	0.007	-0.005	-0.022								
TIM	0.002	**0.024**	0.009	**0.012**	**0.018**	-0.004	0.003	**0.013**	**0.024**	**0.018**	**0.016**	0.007	**0.022**	0.010	-0.001	-0.021	-0.007							
NPT	**0.019**	**0.036**	0.015	**0.023**	**0.023**	-0.003	**0.020**	**0.029**	**0.038**	**0.016**	**0.018**	-0.004	**0.027**	0.011	-0.002	-0.030	0.004	0.008						
STR	0.007	**0.027**	0.013	**0.021**	**0.027**	-0.009	-0.006	**0.019**	**0.026**	0.007	0.002	0.012	**0.017**	-0.010	-0.001	-0.002	0.005	0.006	0.008					
CCB	0.002	0.014	0.004	-0.003	0.014	-0.009	-0.005	0.007	0.010	0.002	0.004	0.005	0.005	0.004	-0.012	-0.012	-0.003	-0.005	0.000	-0.002				
DSD	0.006	0.018	0.006	0.007	0.016	0.000	0.008	0.012	0.012	0.009	0.010	0.014	0.008	0.015	-0.004	-0.022	-0.006	-0.006	0.003	0.009	-0.009			
JCH	0.004	**0.018**	0.006	0.014	0.011	0.002	-0.010	0.010	**0.019**	**0.015**	0.008	0.006	0.016	0.006	0.004	-0.010	0.004	0.001	0.015	0.000	0.000	0.007		

WST	0.010	**0.022**	0.006	0.013	0.007	-0.003	0.001	0.017	**0.023**	0.012	0.007	0.011	0.004	0.000	-0.001	-0.009	0.001	0.001	0.007	-0.001	-0.007	-0.001	-0.008	
OCH	0.014	0.003	0.004	0.011	0.003	**0.030**	0.002	0.008	0.011	0.012	0.015	0.020	0.013	**0.023**	**0.028**	**0.038**	**0.029**	**0.050**	**0.045**	**0.040**	**0.028**	**0.037**	**0.031**	**0.031**

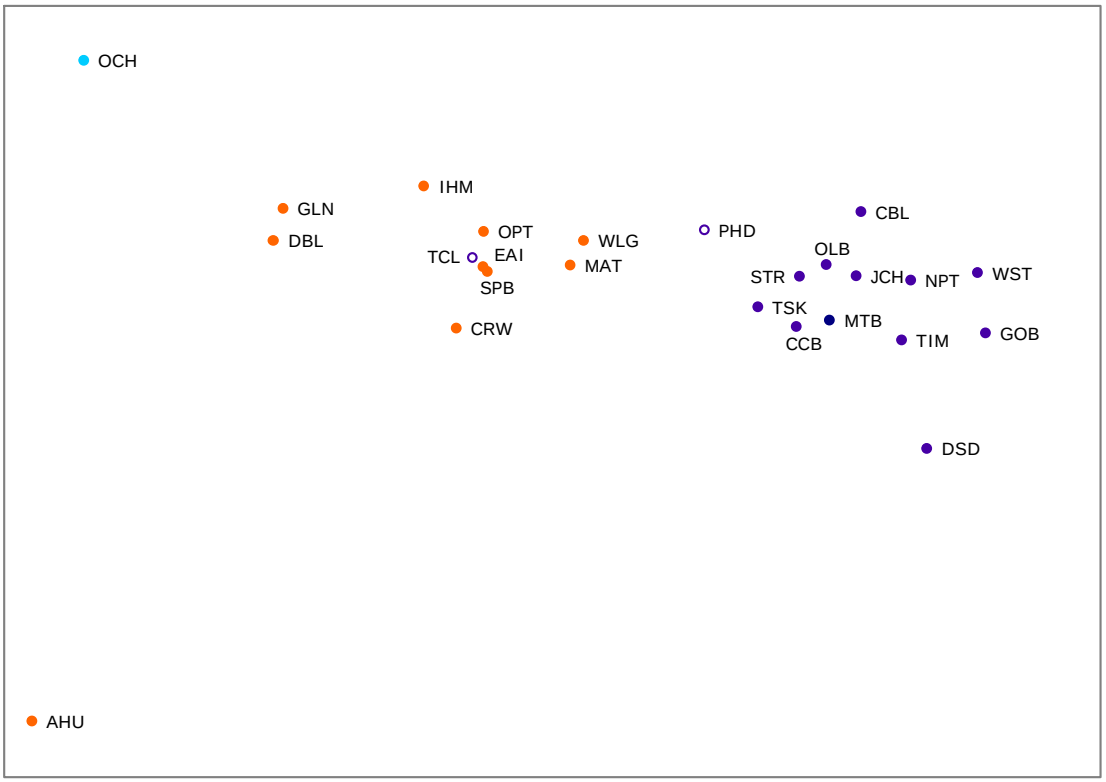


Figure 4.1: Multidimensional scaling based on microsatellites. Nei's genetic distances (d_{xy}) between sampling localities were calculated in Arlequin 3.1 (Excoffier et al. 2005), and MDS was performed in R version 2.6.1 (Team 2007). Stress = 0.582 (calculated according to Venables and Ripley 1999 p. 333). The solid orange dots indicate North Island locations, the solid purple dots indicate South Island locations, the solid blue dot indicates the Chatham Islands, and the white dots outlined in purple indicate the two sampling locations from the north of the South Island (PHD and TCL, Figure 2.3).

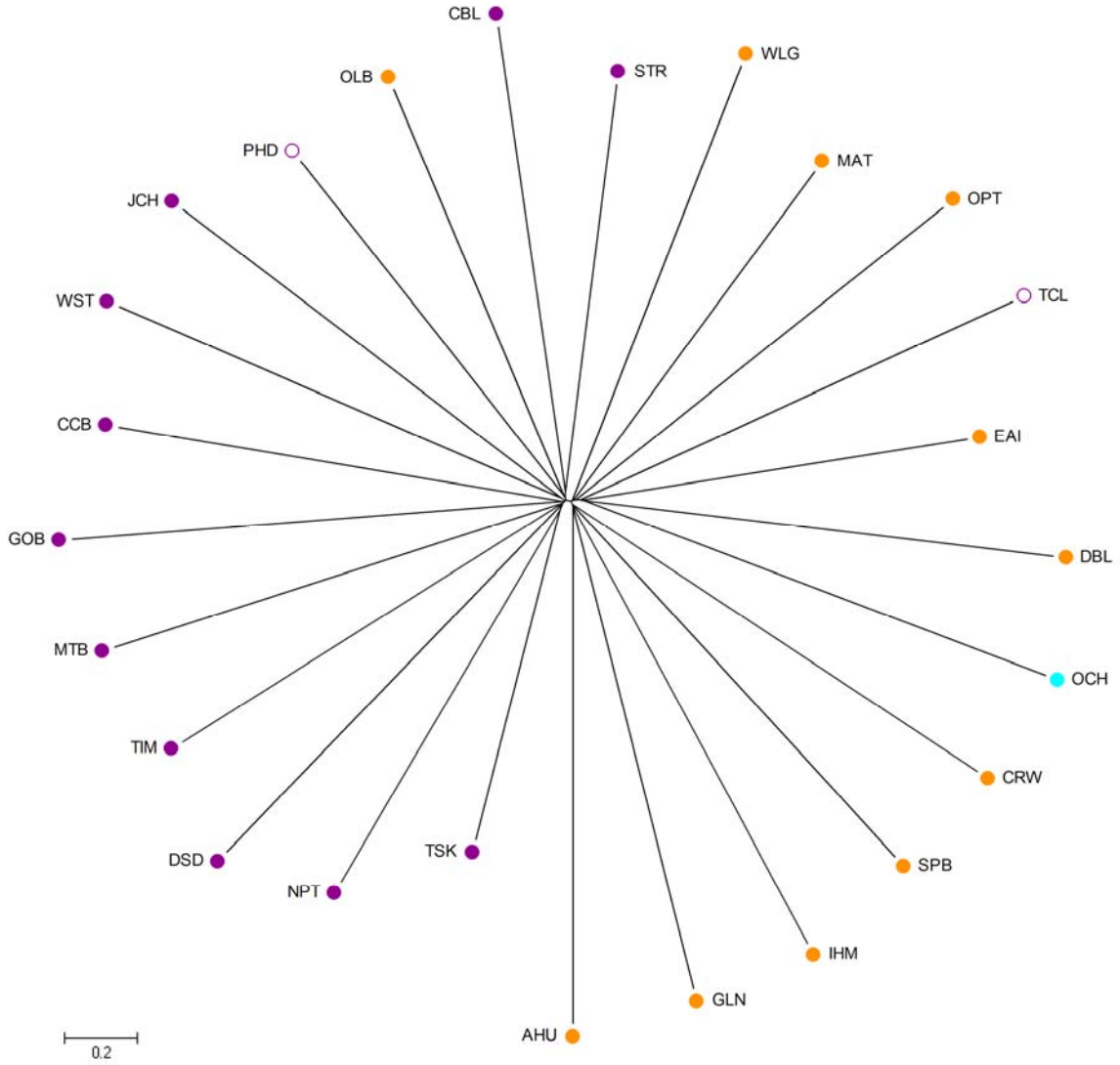


Figure 4.2: Neighbor joining analysis based on microsatellites. Nei's genetic distances (d_{xy}) between sampling localities were calculated in Arlequin 3.1 (Excoffier et al. 2005). The tree was built using the neighbor joining method (Saitou and Nei 1987) in MEGA4 (Tamura et al. 2007). The optimal tree with the sum of branch lengths = 31.650 is shown. The tree is drawn to scale. The solid orange dots indicate North Island locations, the solid purple dots indicate South Island locations, the solid blue dot indicates the Chatham Islands, and the white dots outlined in purple indicate the two sampling locations from the north of the South Island (PHD and TCL, Figure 2.3).

Table 4.3: AMOVA results based on microsatellites. AMOVAs tested the proposed groups listed in Table 2.2. AMOVAs used the weighted average over loci and were calculated in Arlequin 3.1 (Excoffier et al. 2005). * $p < 0.05$, ** $p < 0.01$.

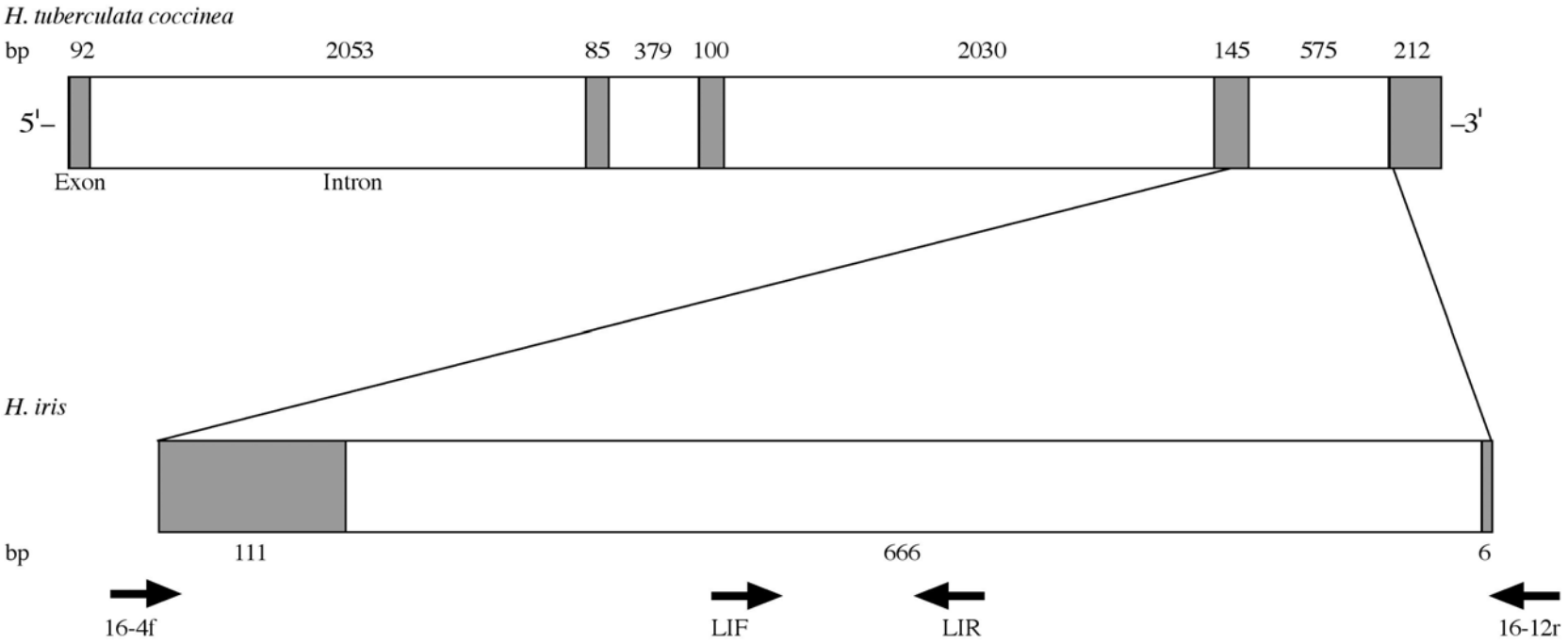
Groups	F_{ST}	F_{SC}	F_{CT}
All locations	0.00900**		
North and South Island Chatham Islands	0.02306**	0.00813**	0.01505*
North Island South Island Chatham Islands	0.01293**	0.00419	0.00878**
North Island and north of South Island Remaining South Island Chatham Islands	0.01367**	0.00340	0.01030**
Excluding OCH			
North and South Island	0.00812**		
North Island South Island	0.01172**	0.00418	0.00757**
North Island and north of South Island Remaining South Island	0.01253**	0.00340	0.00916**

Table 4.4: Standard polymorphism indices for AMOVA groupings based on microsatellites. Listed are N, number of individuals; A, number of alleles; Ae, effective number of alleles; Ar, allelic richness; H_O, observed heterozygosity; and H_E, unbiased expected heterozygosity. All values except allelic richness were calculated in GeneAlEx 6.1 (Peakall and Smouse 2006). Allelic richness was calculated in Fstat 2.9.3 (Goudet 2002).

Group	N	A	Ae	Ar	H_O	H_E
AB14						
All	455	23	7.974	22.928	0.754	0.876
North & South Island	442	23	8.078	8.963	0.758	0.877
North Island	212	20	7.173	8.613	0.816	0.863
South Island	230	19	8.418	9.056	0.704	0.883
North Island & north South Island	248	21	7.129	8.546	0.810	0.861
Remaining South Island	194	19	8.528	9.222	0.691	0.885
Chatham Islands	13	5	3.634	5.000	0.615	0.754
AB21						
All	447	84	35.091	84.000	0.895	0.973
North & South Island	434	83	34.070	19.141	0.892	0.972
North Island	209	73	35.470	19.639	0.885	0.974
South Island	225	59	27.484	18.026	0.898	0.966
North Island & north South Island	245	74	35.943	19.614	0.890	0.974
Remaining South Island	189	55	25.810	17.728	0.894	0.964
Chatham Islands	13	22	18.778	22.000	1.000	0.985
AB31						
All	459	42	6.850	41.838	0.826	0.855
North & South Island	446	41	6.720	10.011	0.823	0.852
North Island	210	32	8.427	10.930	0.848	0.883
South Island	236	29	5.362	8.976	0.801	0.815
North Island & north South Island	247	35	7.991	10.627	0.830	0.877
Remaining South Island	199	27	5.216	8.973	0.814	0.810
Chatham Islands	13	12	9.389	12.000	0.923	0.929

Table 4.5: Mantel tests using microsatellites. Mantel tests were performed within each group proposed in Table 2.2 and implemented in Arlequin 3.1 (Excoffier et al. 2005).

Group	Correlation coefficient (p-value)
All locations	0.233 (0.008)
North & South Islands	0.191 (0.019)
North Island	-0.044 (0.625)
South Island	0.079 (0.242)
North Island & north South Island	-0.152 (0.873)
Remainder South Island	0.104 (0.221)
Chatham Islands	NA



PCR primers		**Internal sequencing primers**	
16-4f	From Metz (1998a)	LIF	5'-CACGCCCACTGTATTGTCAG-3'
16-12r	From Metz (1998a)	LIR	5'-ATTTGCTTTCGGGATACCTT-3'

Figure 5.1: Overview of lysin. Schematic of lysin based on a composite of *H. tuberculata coccinea* sequences from Clark et al. (2007) and Lee et al. (1995). The enlarged region labelled *H. iris* is the fragement examined in this chapter. This fragment was amplified with primers 16-4f and 16-12r and was sequenced with primers 16-4f, 16-12r, LIF, and LIR.

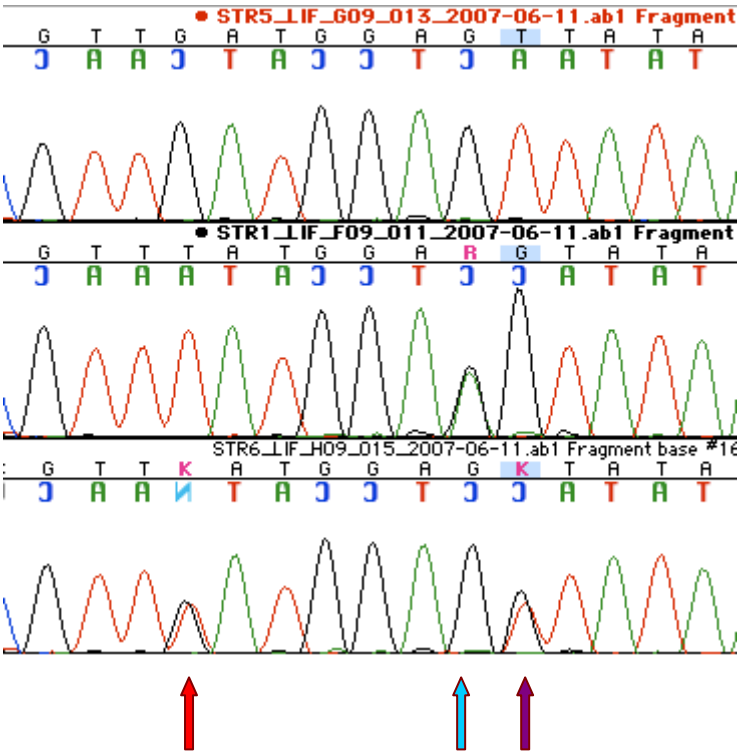


Figure 5.2: Examples of heterozygous base calls for nuclear sequences. Alignment of lysin fragments for three *H. iris* individuals sequenced with primer LIF. STR6 (bottom) is heterozygous (T/G) at the sites indicated with the red and purple arrows. STR1 (middle) is heterozygous (A/G) at the site indicated with the blue arrow. Heterozygous base calls were confirmed with sequencing in the reverse direction. Image was produced in Sequencher™ 4.2.2 (Gene Codes Corporation).

Table 5.1: Comparison of the amount of differentiation between the samples used in Chapters 2 and 4 and those used in Chapter 5. Φ_{ST} based on mitochondrial DNA and F_{ST} based on microsatellites are presented for the 25 samples used in Chapter 2 and 4 and the 17 samples used in Chapter 5. Indices were calculated in Arelequin 3.1 (Excoffier et al. 2005).

	Φ_{ST} (p-value)	F_{ST} (p-value)
25 samples	0.04453 (0.000)	0.00900 (0.000)
Subset of 17 samples	0.05705 (0.000)	0.00859 (0.001)
Excluding OCH		
24 samples	0.03815 (0.000).	0.00812 (0.000)
Subset of 16 samples	0.04918 (0.000)	0.00695 (0.009)

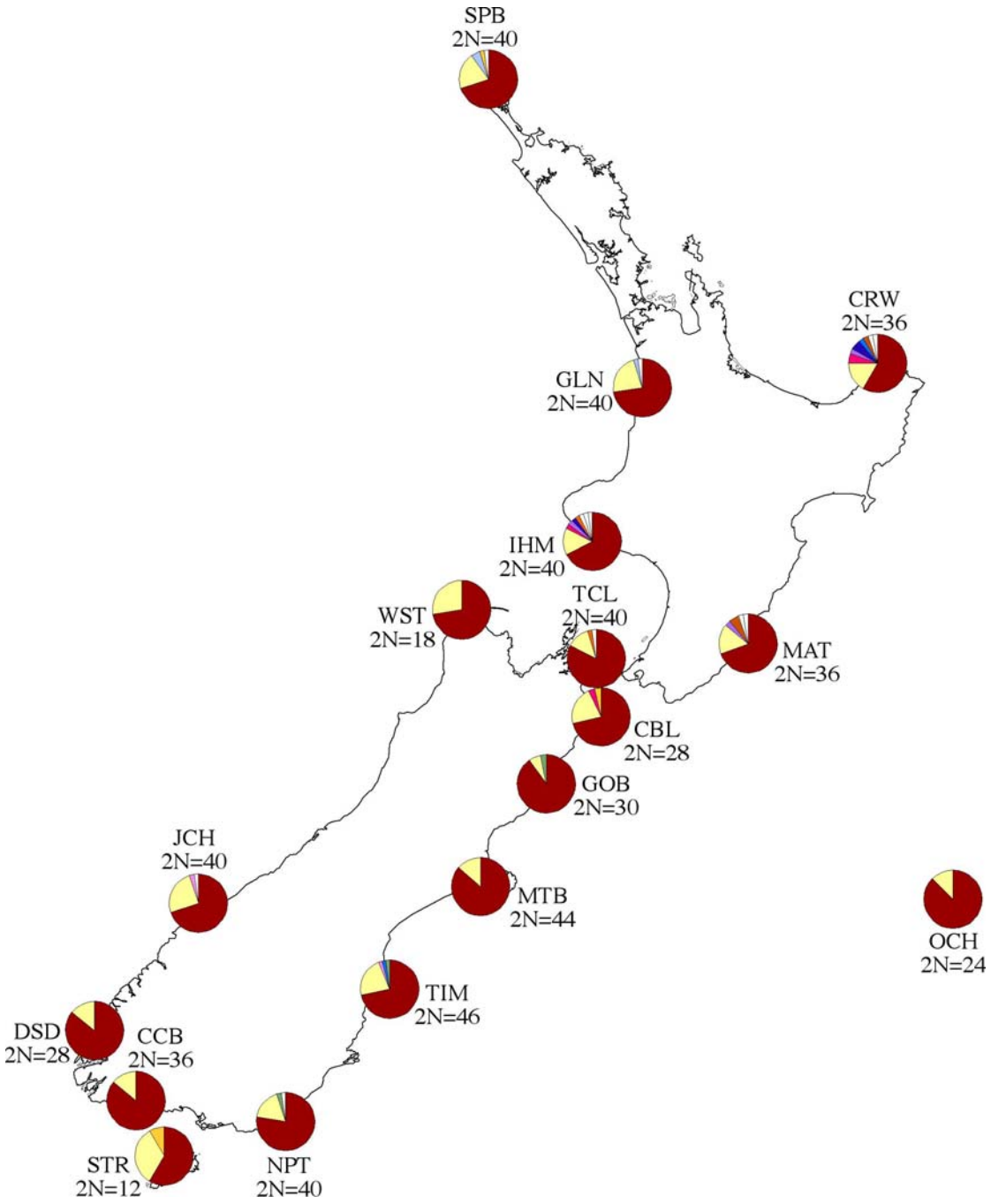


Figure 5.3: Lysin fragment haplotype frequencies and sample sizes at collection locations around New Zealand. Haplotypes were inferred using PHASE2.1.1 (Stephens et al. 2001; Stephens and Donnelly 2003). Colored haplotypes are shared among sampling locations, while white haplotypes are private. Colors correspond to haplotypes in Figure 5.5. Haplotype sequences and genotype frequencies are listed in Appendices 8 and 9, respectively.

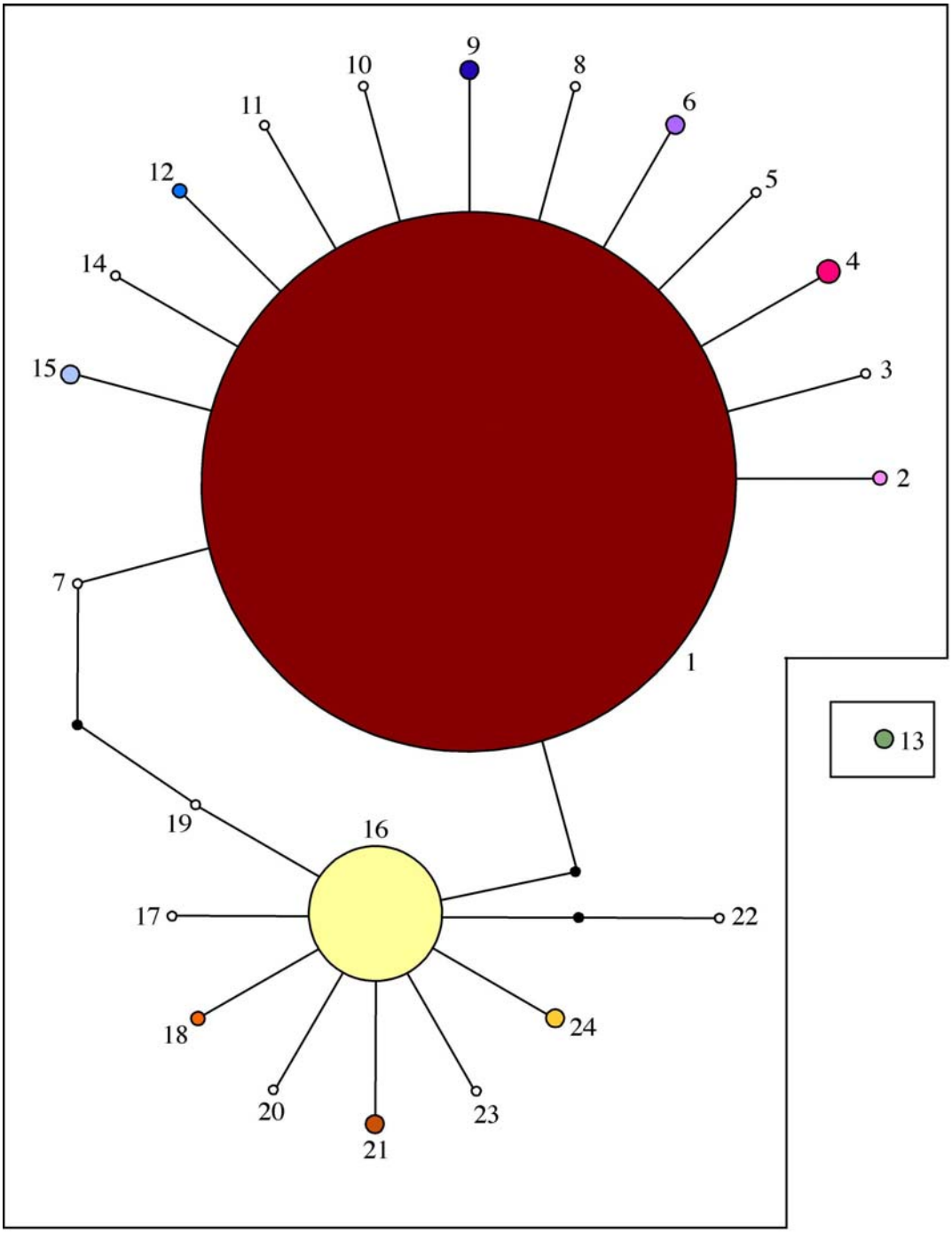


Figure 5.4: Statistical parsimony networks based on lysin fragments. TCS network of lysin haplotypes. Relationships among haplotypes were inferred using statistical parsimony (Templeton et al. 1992), implemented in TCS (Clement et al. 2000). Colored haplotypes are shared among sampling locations, white haplotypes are private, and black haplotypes are missing haplotypes. Haplotypes are labeled from 1–24, and haplotype sequences are listed in Appendix 8.

Table 5.2: Polymorphism data and neutrality results for 783 bp of lysin across 17 locations around New Zealand. Individuals are grouped according to sampling locations (Figure 2.3). Results were generated in Arlequin 3.1 (Excoffier et al. 2005).

ID	2N	Number of polymorphic sites	Number of haplotypes	H_O (p-value)	H_E±S.D.	Nucleotide diversity (π±S.D.)	Tajima's *D* (p-value)	Fu's F_s (p-value)
CBL	28	4	4	0.4286 (1.000)	0.4577±0.0960	0.001176±0.000932	-0.27301 (0.425)	0.13701 (0.530)
CCB	36	2	2	0.2778 (1.000)	0.2460±0.0838	0.000628±0.000613	0.03933 (0.643)	1.71548 (0.738)
CRW	36	10	9	0.7222 (0.838)	0.6397±0.0832	0.001610±0.001155	-1.45422 (0.057)	-3.73693 (0.016)
DSD	28	2	2	0.2857 (1.000)	0.2540±0.0953	0.000649±0.000630	-0.02446 (0.437)	1.62955 (0.733)
GLN	40	4	4	0.4500 (1.000)	0.4333±0.0779	0.001041±0.000848	-0.31736 (0.421)	0.15860 (0.532)
GOB	30	46	3	0.2000 (1.000)	0.1908±0.0928	0.004075±0.002411	-1.00369 (0.188)	6.31434 (0.984)
IHM	40	9	9	0.4500 (0.286)	0.5308±0.0891	0.001361±0.001021	-1.31237 (0.091)	-4.30584 (0.006)
JCH	40	4	4	0.5000 (1.000)	0.4577±0.0734	0.001110±0.000886	-0.18066 (0.464)	0.31681 (0.562)
MAT	36	6	6	0.6111 (1.000)	0.4984±0.0916	0.001336±0.001011	-0.75702 (0.249)	-1.27873 (0.221)
MTB	44	2	2	0.1818 (0.323)	0.2410±0.0756	0.000616±0.000602	0.08917 (0.639)	1.78783 (0.750)
NPT	40	47	4	0.4500 (1.000)	0.3769±0.0848	0.003630±0.002170	-0.19467 (0.443)	4.56412 (0.959)
OCH	24	2	2	0.2500 (1.000)	0.2283±0.1021	0.000583±0.000594	-0.32459 (0.350)	1.33315 (0.653)
SPB	40	5	5	0.5000 (1.000)	0.4782±0.0823	0.001235±0.000954	-0.45205 (0.374)	-0.42138 (0.401)
STR	12	3	3	0.6667 (0.635)	0.5909±0.1079	0.001567±0.001203	0.77220 (0.793)	1.11706 (0.736)
TCL	40	4	4	0.3500 (1.000)	0.3103±0.0882	0.000796±0.000710	-0.40864 (0.372)	-0.48167 (0.346)
TIM	46	48	5	0.3913 (0.807)	0.4464±0.0743	0.003443±0.002070	-0.32386 (0.415)	3.23017 (0.906)
WST	18	2	2	0.5556 (1.000)	0.4248±0.0993	0.001085±0.000901	1.12564 (0.867)	2.42347 (0.858)

Table 6.1: Gα1 intron primers. G5F, G5R, G6F, and G6R were designed using Primer3 (Rozen and Skaletsky 2000). PCR reactions used G5F and S2. Sequencing reactions used either G5F, G5R, G6F, or G6R.

Primer name	Sequence (5'–3')
S1	See (Wodicka and Morse 1991)
S2	See (Wodicka and Morse 1991)
G5F	CGTCCATCATGTTCTTAGTAGCC
G5R	CAAGTCCGGTCCAGACAATC
G6F	AAAAGGAAGAGGCGACTAAGG
G6R	AGCCCCTTAATACCGAGTGC

Table 6.2: Comparison of the amount of differentiation between the samples used in Chapters 2, 4, and 5 and those used in Chapter 6. Φ_{ST} based on mitochondrial DNA and F_{ST} based on microsatellites are presented for the 25 samples used in Chapter 2 and 4, the 17 samples used in Chapter 5, and the 14 samples used in Chapter 6. Indices were calculated in Arelequin 3.1 (Excoffier et al. 2005).

	Φ_{ST} (p-value)	F_{ST} (p-value)
25 samples	0.04453 (0.000)	0.00900 (0.000)
Subset of 17 samples	0.05705 (0.000)	0.00859 (0.001)
Subset of 14 samples	0.05094 (0.000)	0.00759 (0.004)
Excluding OCH		
24 samples	0.03815 (0.000).	0.00812 (0.000)
Subset of 16 samples	0.04918 (0.000)	0.00695 (0.009)
Subset of 13 samples	0.03985 (0.001	0.00557 (0.024)

Table 6.3: Polymorphism data and neutrality results for 857 bp of Gα1 intron across 14 locations around New Zealand. Individuals are grouped according to sampling locations (Figure 2.3). Results were generated in Arlequin 3.1 (Excoffier et al. 2005).

ID	2N	Number of polymorphic sites	Number of haplotypes	H_O (p-value)	$H_E \pm$ S.D.	Nucleotide diversity ($\pi \pm$S.D.)	Tajima's *D* (p-value)	Fu's F_s (p-value)
CCB	34	31	19	0.7647 (0.372)	0.9269±0.0283	0.007174±0.003889	-1.51897 (0.047)	-5.19180 (0.035)
CRW	26	41	19	0.9231 (0.297)	0.9723±0.0183	0.008308±0.004486	-1.29791 (0.087)	-6.89249 (0.008)
DSD	22	38	16	0.9091 (0.566)	0.9610±0.0278	0.010121±0.005422	-0.81902 (0.225)	-3.66001 (0.071)
GLN	38	33	24	1.0000 (1.000)	0.9701±0.0129	0.007505±0.004038	-1.21198 (0.100)	-9.87953 (0.003)
GOB	36	38	19	0.8889 (0.729)	0.9222±0.0313	0.007788±0.004182	-1.23694 (0.094)	-4.15022 (0.068)
IHM	40	84	31	1.0000 (1.000)	0.9833±0.0105	0.012350±0.006384	-1.68647 (0.024)	-13.61938 (0.001)
MAT	32	51	24	0.9375 (0.512)	0.9778±0.0147	0.010418±0.005484	-1.02852 (0.156)	-9.12274 (0.005)
NPT	40	37	17	0.8000 (0.423)	0.8808±0.0421	0.007814±0.004182	-1.43028 (0.055)	-1.93188 (0.252)
OCH	26	40	16	0.9231 (0.324)	0.9477±0.0267	0.011836±0.006225	-0.98564 (0.170)	-1.60648 (0.275)
SPB	42	67	31	1.0000 (0.802)	0.9826±0.0097	0.011576±0.006001	-1.49081 (0.044)	-13.26983 (0.001)
STR	12	21	9	1.0000 (1.000)	0.9394±0.0577	0.008081±0.004605	-0.84805 (0.207)	-1.26871 (0.233)
TCL	40	36	19	0.9500 (0.988)	0.9192±0.0265	0.007542±0.004050	-1.50236 (0.049)	-3.64600 (0.107)
TIM	46	36	18	0.9565 (0.550)	0.9227±0.0222	0.009651±0.005057	-0.71009 (0.261)	-0.86969 (0.408)
WST	20	33	14	1.0000 (1.000)	0.9579±0.0281	0.009351±0.005065	-1.33578 (0.076)	-2.69149 (0.123)

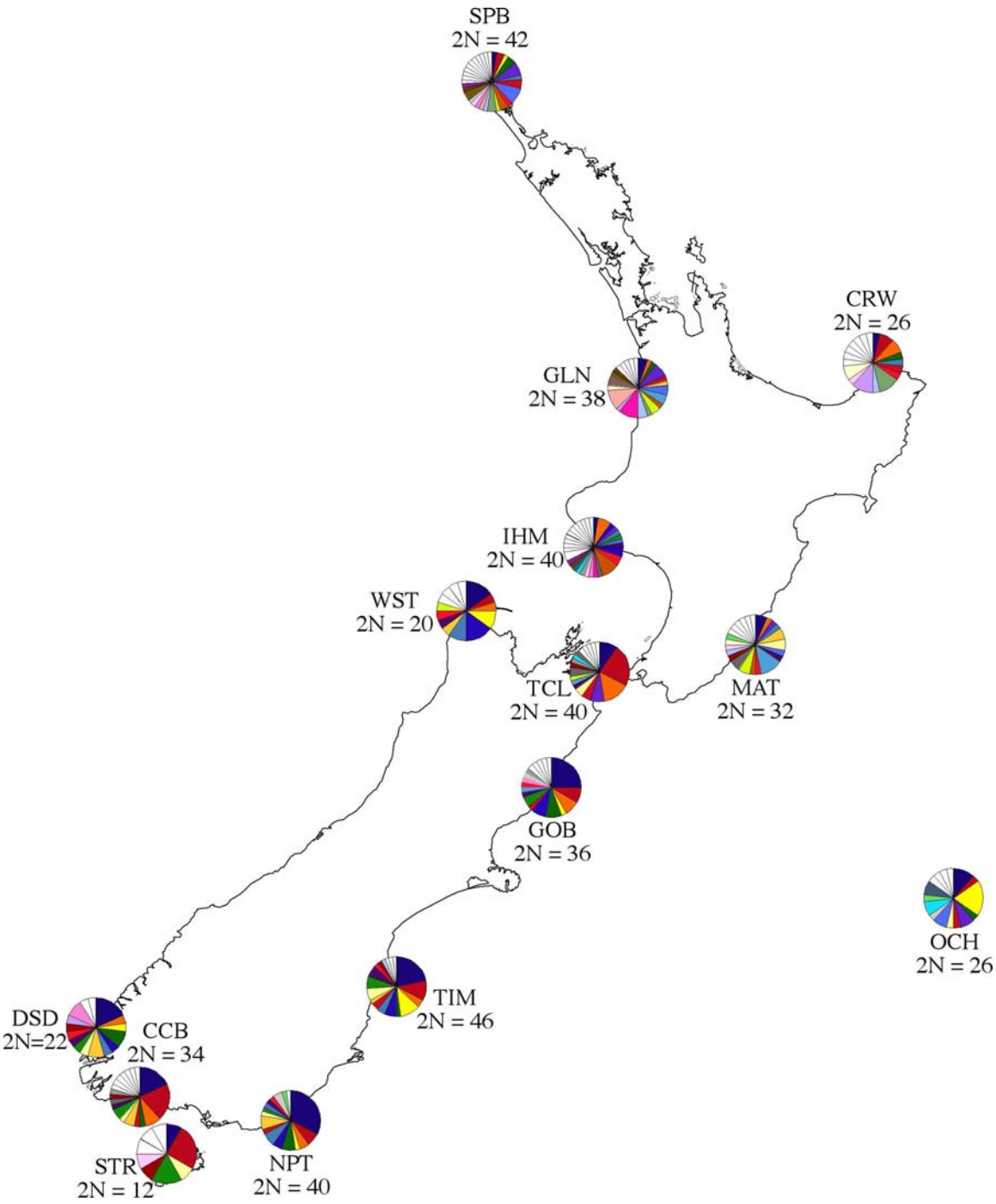


Figure 6.1: Gα1 intron haplotype frequencies and sample sizes at 14 locations around New Zealand. Colored haplotypes are shared aamong sampling locations, while white haplotypes are private. Colors correspond to haplotypes in Figure 6.2. Haplotype sequences and genotype frequencies are listed in Appendices 10 and 11, respectively.

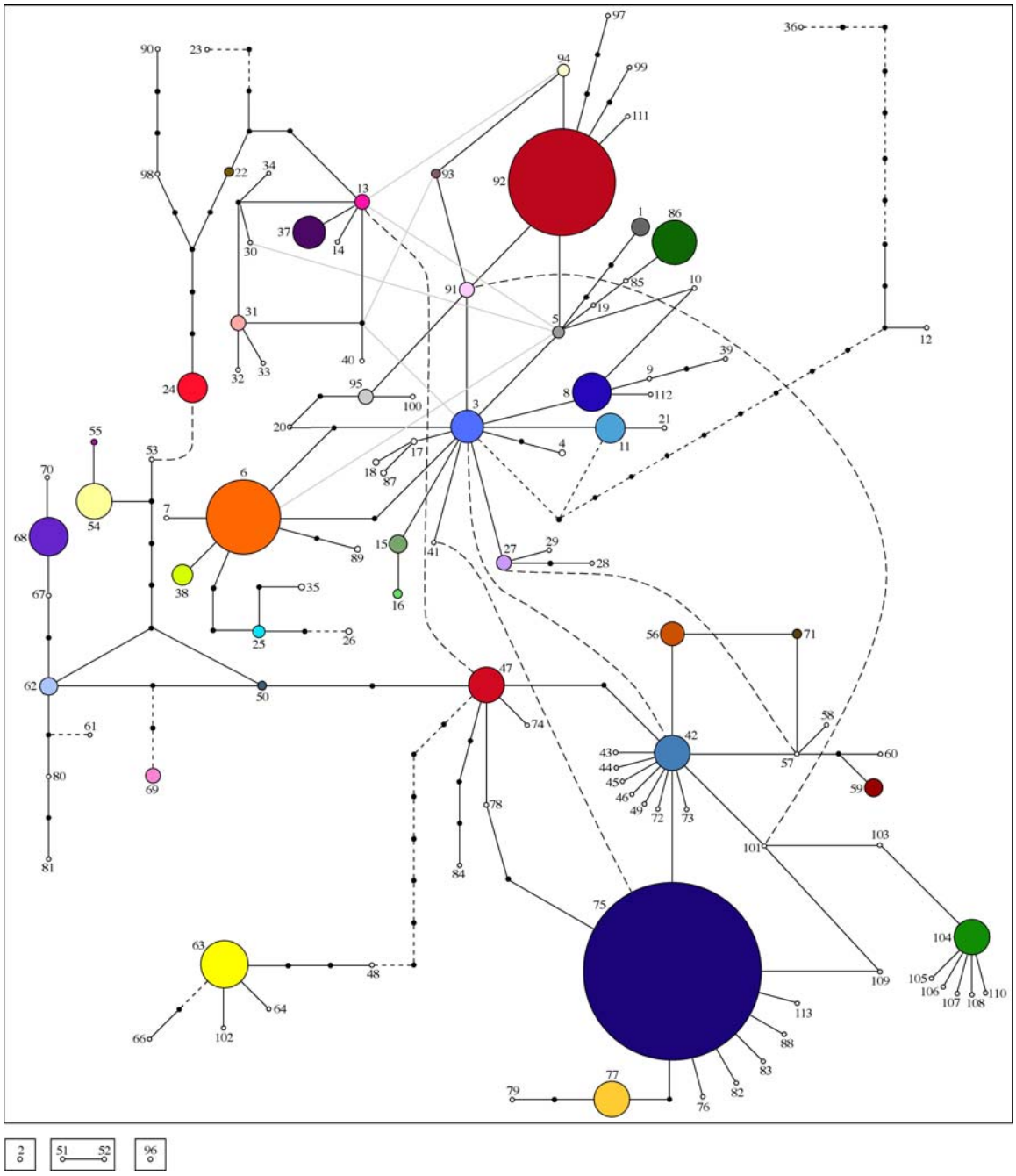


Figure 6.2: Statistical parsimony networks based on Gα1 introns. Relationships among haplotypes were inferred using statistical parsimony (Templeton et al. 1992), implemented in TCS (Clement et al. 2000). Colored haplotypes are shared among sampling locations, white haplotypes are private, and black haplotypes are missing haplotypes. Solid black and gray lines represent a one base pair change. Lines that overlap other lines are colored gray for clarity. Small dashed lines between nodes represent a base pair indel. Large dashed lines between nodes represent a 5 bp indel. Haplotypes are labeled from 1–113 (there is no haplotype 65), and haplotype sequences are listed in Appendix 10.

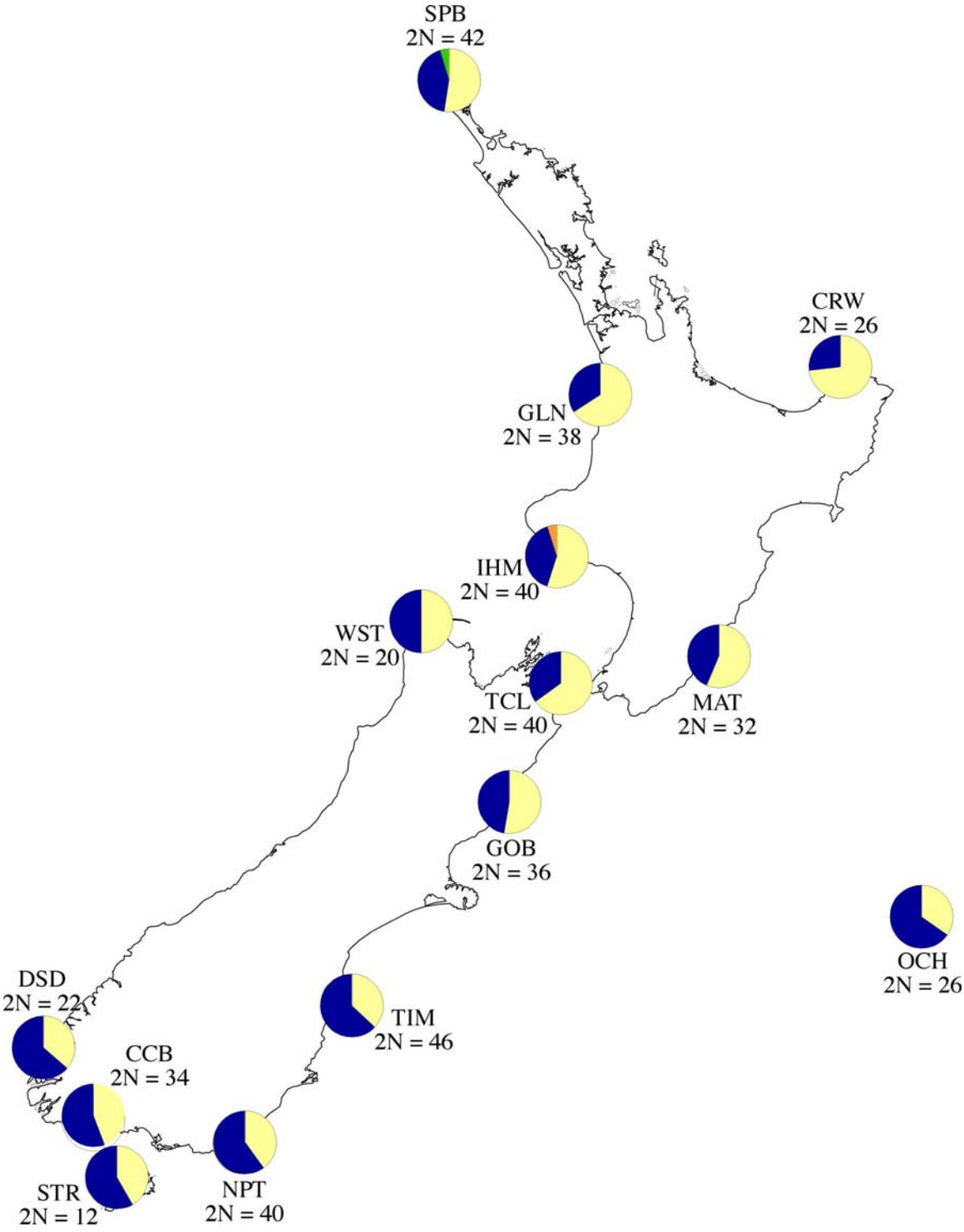


Figure 6.3: Frequency of a 5 bp deletion at 14 locations around New Zealand. Blue represents haplotypes with a 5 bp insertion. Yellow represents haplotypes with a 5 bp deletion. Green represents haplotypes 51 and 52, which have a 5 bp insertion, but were not connected to the main network (Figure 6.2). Orange represents haplotypes 2 and 96, which have a 5 bp deletion, but were not connected to the main network (Figure 6.2).

Table 6.4: Pairwise Φ_{ST} based on Gα1 intron. Pairwise Φ_{ST} indices were calculated in Arelequin 3.1 (Excoffier et al. 2005). Significances were tested with 16002 permutations. Bold blue text indicates $p < 0.05$. No samples were significant after Bonferroni corrections ($p < 0.0006$). Φ_{ST} = 0.000 was used for $\Phi_{ST} < 0.0005$. Orange samples with black text are North Island locations. The blue sample with black text is the Chatham Islands. Purple samples with white text are South Island locations. White samples with purple text are from the north of the South Island (Figure 2.3).

	SPB	CRW	MAT	GLN	IHM	TCL	GOB	TIM	NPT	STR	CCB	DSD	WST
CRW	0.007												
MAT	-0.005	0.013											
GLN	0.003	0.000	0.006										
IHM	-0.013	-0.001	-0.010	-0.001									
TCL	0.022	0.000	0.020	0.014	0.006								
GOB	0.011	0.021	0.010	0.031	0.012	0.017							
TIM	0.016	**0.070**	0.022	**0.062**	**0.033**	**0.065**	0.016						
NPT	**0.033**	**0.074**	0.026	**0.075**	**0.042**	**0.070**	-0.008	0.007					
STR	0.033	**0.089**	0.049	**0.103**	0.042	**0.071**	0.026	0.027	0.043				
CCB	**0.035**	**0.067**	0.031	**0.055**	0.021	**0.039**	-0.011	0.015	-0.008	0.011			
DSD	0.003	**0.065**	0.005	**0.070**	**0.038**	**0.067**	0.004	-0.022	-0.007	0.019	0.006		
WST	-0.006	0.017	-0.008	0.021	-0.004	0.022	-0.018	-0.013	-0.013	0.047	0.003	-0.017	
OCH	0.029	**0.095**	**0.054**	**0.087**	**0.051**	**0.105**	**0.081**	0.008	**0.080**	**0.082**	**0.092**	0.011	0.025

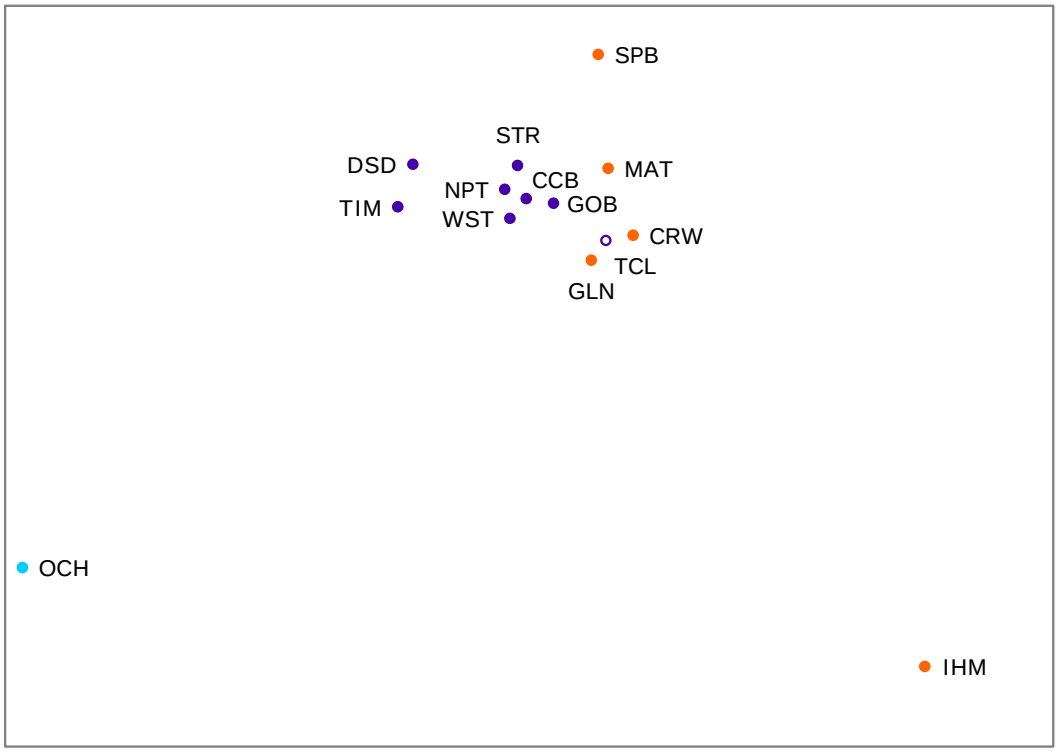


Figure 6.4: Multidimensional scaling based on Gα1 intron. Nei's genetic distances (d_{xy}) between sampling localities were calculated in Arlequin 3.1 (Excoffier et al. 2005). MDS was performed in R version 2.6.1 (Team 2007). Stress = 0.55 (calculated according to Venables and Ripley 1999 p. 333). The solid orange dots indicate North Island locations, the solid purple dots indicate South Island locations, the solid blue dot indicates the Chatham Islands, and the white dots outlined in purple indicate the two sampling locations from the north of the South Island (PHD and TCL, Figure 2.3).

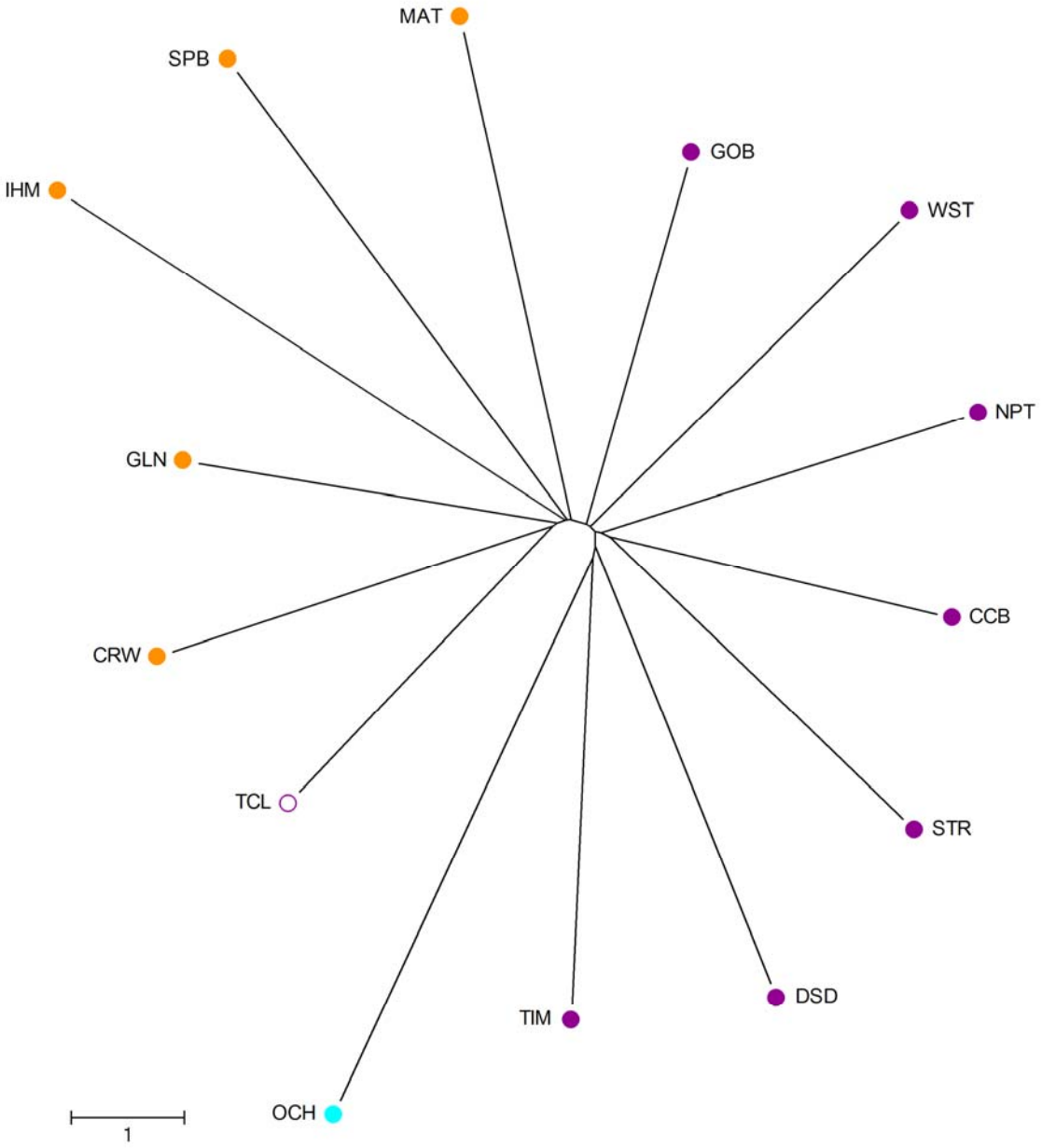


Figure 6.5: Neighbor joining analysis based on Gα1 intron. Nei's genetic distances (d_{xy}) between sampling localities were calculated in Arlequin 3.1 (Excoffier et al. 2005). The tree was built using the neighbor joining method (Saitou and Nei 1987) in MEGA4 (Tamura et al. 2007). The optimal tree with the sum of branch lengths = 55.704 is shown. The tree is drawn to scale. The solid orange dots indicate North Island locations, the solid purple dots indicate South Island locations, the solid blue dot indicates the Chatham Islands, and the white dots outlined in purple indicate the two sampling locations from the north of the South Island (PHD and TCL, Figure 2.3).

Table 6.5: AMOVA results based on Gα1 introns. AMOVAs tested the proposed groups listed in Table 2.2. AMOVAs were based on pairwise distances and implemented in Arelquin 3.1 (Excoffier et al. 2005). *p < 0.05, **p < 0.01.

Groups	Φ_{ST}	Φ_{SC}	Φ_{CT}
All locations	0.02666**		
North and South Island Chatham Islands	0.06993**	0.02062**	0.05034
North Island South Island Chatham Islands	0.03954**	0.00791	0.03188**
North Island and north of South Island Remaining South Island Chatham Islands	0.04253**	0.00200	0.04062**
Excluding OCH			
North and South Island	0.02146**		
North Island South Island	0.03262**	0.00855	0.02428**
North Island and north of South Island Remaining South Island	0.03668**	0.00255	0.03422**

Table 6.6: Polymorphism data and neutrality results for Gα1 intron sequences across AMOVA groupings. Locations within groups were specified in Table 2.2. Results generated in Arelequin 3.1 (Excoffier et al. 2005).

Group	2N	Number of polymorphic sites	Number of haplotypes	H_O	$H_E \pm$ S.D.	Nucleotide diversity π±S.D.	Tajima's *D* (p-value)	Fu's F_s (p-value)
All locations	454	129	112	0.9295 (0.441)	0.9626±0.0040	0.009493±0.004881	-1.86033 (0.005)	-24.09380 (0.002)
North & South Islands	428	128	107	0.9299 (0.830)	0.9619±0.0042	0.009288±0.004784	-1.87861 (0.003)	-24.14520 (0.002)
North Island	178	116	77	0.9775 (0.876)	0.9822±0.0024	0.010170±0.005220	-1.86949 (0.006)	-24.40713 (0.000)
South Island	250	64	58	0.8960 (0.653)	0.9279±0.0091	0.008479±0.004409	-1.48507 (0.036)	-22.70422 (0.001)
North Island & north South Island	218	119	82	0.9725 (0.883)	0.9779±0.0028	0.009721±0.005001	-1.87374 (0.005)	-24.37388 (0.000)
Remainder South Island	210	54	50	0.8857 (0.591)	0.9232±0.0109	0.008524±0.004434	-1.19881 (0.093)	-16.17059 (0.004)
Chatham Islands	26	40	16	0.9231 (0.345)	0.9477±0.0267	0.011836±0.006225	-0.98564 (0.163)	-1.60648 (0.268)

Table 6.7: Mantel tests using Gα1 introns. Mantel tests were performed within each group proposed in Table 2.2 and implemented in Arlequin 3.1 (Excoffier et al. 2005).

Group	Correlation coefficient (p-values)
All locations	0.493937 (0.000)
North & South Islands	0.487007 (0.001)
North Island	0.118051 (0.361)
South Island	0.317915 (0.100)
North Island & north South Island	-0.133121 (0.680)
Remainder South Island	0.266717 (0.185)
Chatham Islands	NA

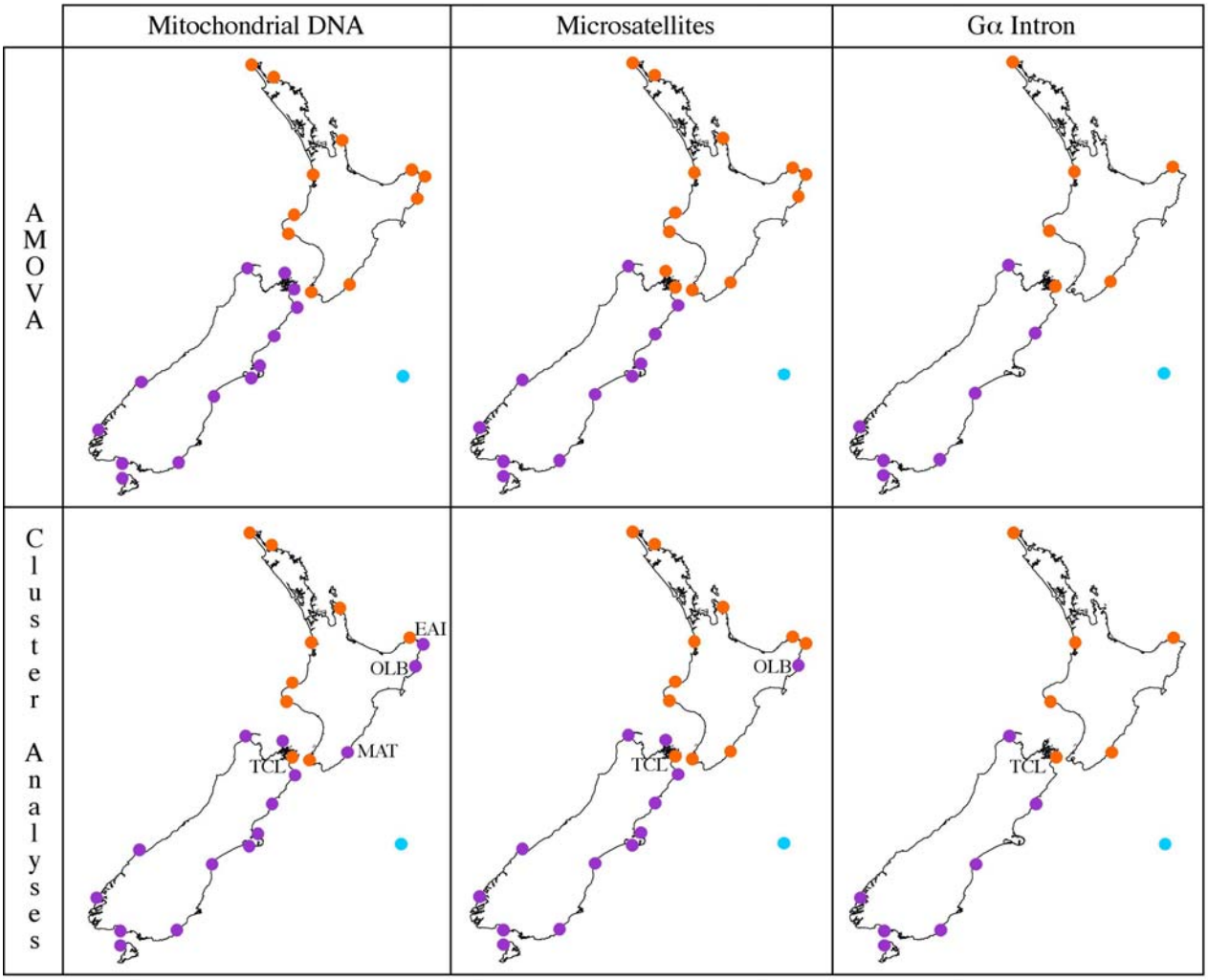


Figure 7.1: Comparison of AMOVA and cluster analyses. Image depicts a summary of results from Chapter 2 (mitochondrial DNA), Chapter 4 (microsatellites), and Chapter 6 (Gα1 intron). Lysin is not pictured because no significant genetic structure was detected using this locus.

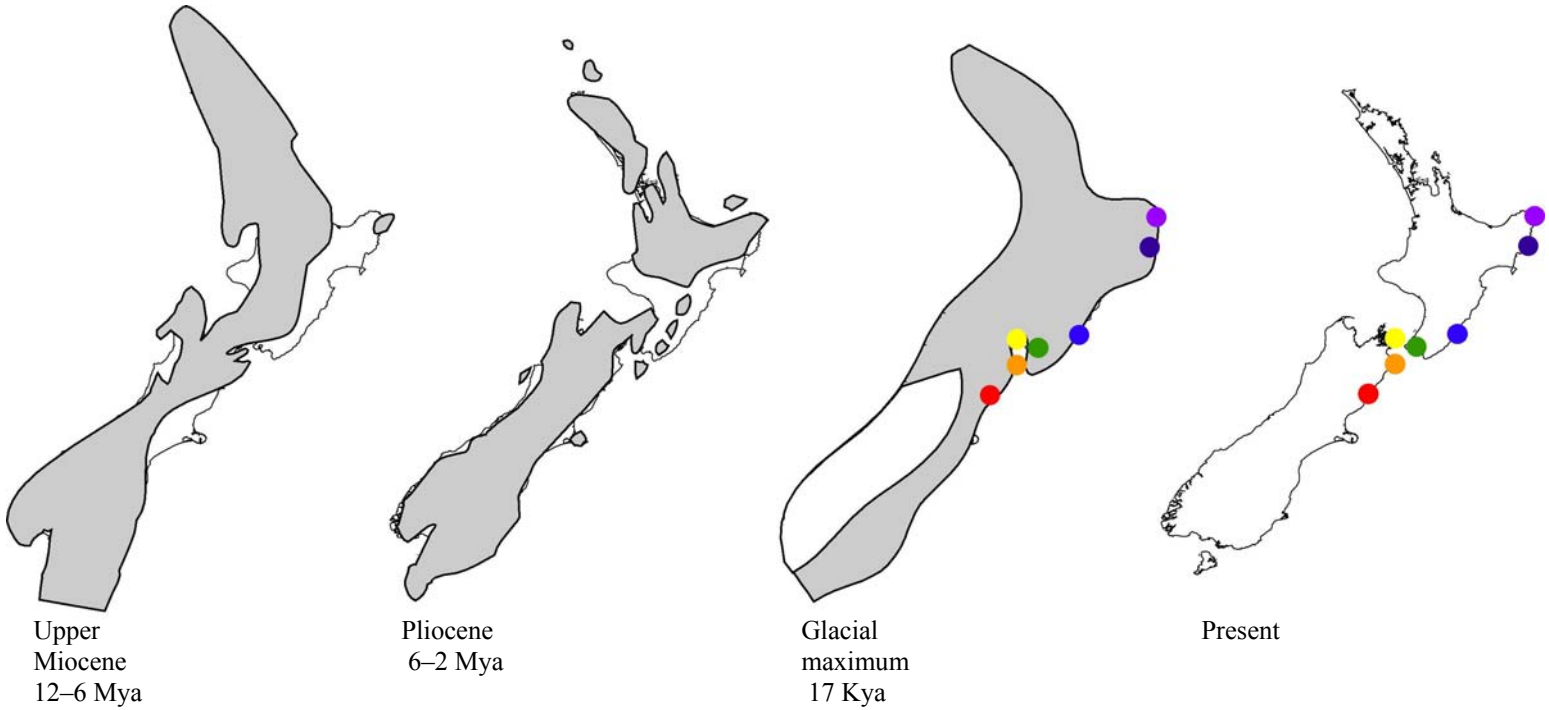


Figure 7.2: Reconstruction of the New Zealand land mass over the last 12 Myr. Approximate locations for samples GOB (red), CBL (orange), TCL (yellow), WLG (green), MAT (light blue), OLB (dark blue), and EAI (purple) have been superimposed on the 17 Kya image and the present image. Gray represents land that was above sea level during the time period indicated (Stevens and Hogg 2004). The changing landmass of New Zealand is from Figure 1 in Stevens and Hogg (2004), originally adapted from from Fleming (1979) and Stevens et al. (1995)

Acknowledgements

Thank you to my supervisors Drs. Neil Gemmell and Marie Hale for their time and patience. This project relied on samples of abalone collected from all around New Zealand. Thank you to the many who have collected samples: Ed Arron, Paul Buisson (Department of Conservation), Christine Conway (Burkheart Fisheries), Paul Ferguson, Peter Keesing, Erica Mendieta, Peter Molloy, Don Neale (Department of Conservation), Gerard Prendeville, Chris Squire & Crew (Riverton Fisherman, Ltd), Tammy Steeves, Ham Stephen & Crew (Ministry of Fisheries), Brian Williams (Deparment of Conservation), and Martin Williams & Crew (Ministry of Fisheries). I have had lots of help collecting these samples, and I apologize in advance if I have missed someone vital to these collections.

Thank you to Dr. Sharyn Goldstien for her help, guidance, and support concerning various aspects of this project and Dr. Maxine Bryant for technical assistance with the writing of this thesis and for food and shelter for the many months of writing the thesis. Thank you to Aleksandr Kalinin for help screening the microsatellites in Chapters 3 and 4. To Thorsten and Iris, I valued your encouragement and the sustenance you provided over the past couple of years. For their assistance throughout the duration of the project I am grateful to the many members of the Molecular Ecology Laboratory: Andrew Bagshaw, Manu Buschiazzo, Antoine Fouquet, Erica Mendieta, Kathrin McBride, Angelika Merkel, Genievive Manalo del Mundo, Melanie Pierson, and Tammy Steeves. I was extremely lucky to have also had such a great support crew from my friends in New Zealand and overseas.

I am forever grateful to Snadra Negro who has literally been here since the commencement of this project to the very last of the corrections. She has offered an endless amount of encouragement, home cooked meals, and many diversions away from the thesis. The years of this thesis have been far from pleasant, but I survived thanks to you!

I have also enjoyed the eternal love, support, and patience of my family as I have studied overseas. Thank you!

This project was possible with funding from the University of Canterbury, Sigma Xi, Ministry of Fisheries, Education New Zealand, MIRHT, and the Royal Society of New Zealand, Canterbury Branch.

Literature Cited

Addison, J. A., and M. W. Hart. 2004. Analysis of population genetic structure of the green sea urchin (*Strongylocentrotus droebachiensis*) using microsatellites. *Marine Biology* 144: 243-251.

Adkins, R. M. 2004. Comparison of the accuracy of methods of computational haplotype inference using a large empirical dataset. *BMC Genetics* 5: 22.

Ahmed, F., Y. Koike, C. A. Strussmann, I. Yamasaki, M. Yokota, and S. Watanabe. 2008. Genetic characterization and gonad development of artificially produced interspecific hybrids of the abalones, *Haliotis discus discus* Reeve, *Haliotis gigantea* Gmelin and *Haliotis madaka* Habe. *Aquaculture Research* 39: 532-541.

Aliani, S., and A. Molcard. 2003. Hitch-hiking on floating marine debris: macrobenthic species in the Western Mediterranean Sea. *Hydrobiologia* 503: 59-67.

Allen, V. J., I. D. Marsden, N. L. C. Ragg, and S. Gieseg. 2006. The effects of tactile stimulants on feeding, growth, behaviour, and meat quality of cultured Blackfoot abalone, *Haliotis iris*. *Aquaculture* 257: 294-308.

Allendorf, F. W., P. R. England, G. Luikart, P. A. Ritchie, and N. Ryman. 2008. Genetic effects of harvest on wild animal populations. *Trends in Ecology and Evolution* 23: 327-337.

Amos, W., J. I. Hoffman, A. Frodsham, L. Zhang, S. Best, and A. V. S. Hill. 2007. Automated binning of microsatellite alleles: Problems and solutions. *Molecular Ecology Notes* 7: 10-14.

An, H. S., and S. J. Han. 2006. Isolation and characterization of microsatellite DNA markers in the Pacific abalone, *Haliotis discus hannai*. *Molecular Ecology Notes* 6: 11-13.

Anderson, F. M. 1902. Cretaceous deposits of the Pacific coast. *Proceedings of the California Academy of Sciences* 2: 1-154.

Andrés, A. M., A. G. Clark, L. Shimmin, E. Boerwinkle, C. F. Sing, and J. E. Hixson. 2007. Understanding the accuracy of statistical haplotype inference with sequence data of known phase. *Genetic Epidemiology* 31: 659-671.

Apte, S., and J. P. A. Gardner. 2001. Absence of population genetic differentiation in the New Zealand greenshell mussel *Perna canaliculus* (Gmelin 1791) as assessed by allozyme variation. *Journal of Experimental Marine Biology and Ecology* 258: 173-194.

Apte, S., and J. P. A. Gardner. 2002. Population genetic subdivision in the New Zealand greenshell mussel (*Perna canaliculus*) inferred from single-strand conformation polymorphism analysis of mitochondrial DNA. *Molecular Ecology* 11: 1617-1628.

Apte, S., B. Star, and J. P. A. Gardner. 2003. A comparison of genetic diversity between cultured and wild populations, and a test for genetic introgression in the New Zealand greenshell mussel Perna canaliculus (Gmelin 1791). *Aquaculture* 219: 193-220.

Avise, J. C. 2004. *Molecular Markers, Natural History, and Evolution*. Sinauer Associates, Inc. Publishers, Sunderland, MA USA.

Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb, and N. C. Saunders. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual review of ecology and systematics.* 18: 489-522.

Ayers, K. L., and J. M. Waters. 2005. Marine biogeographic disjunction in central New Zealand. *Marine Biology* 147: 1045-1052.

Babcock, R., and J. Keesing. 1999. Fertilization biology of the abalone *Haliotis laevigata*: laboratory and field studies. *Canadian Journal of Fisheries and Aquatic Sciences* 56: 1668-1678.

Backström, N., S. Fagerberg, and H. Ellegren. 2008. Genomics of natural bird populations: A gene-based set of reference markers evenly spread across the avian genome. *Molecular Ecology* 17: 964-980.
Baker, A. J. e. 2000. *Molecular Methods in Ecology*. Blackwell Science Ltd., London.
Ballard, J. W. O., and M. C. Whitlock. 2004. The incomplete natural history of mitochondria. *Molecular Ecology* 13: 729-744.
Balloux, F., H. Brunner, N. Lugon-Moulin, J. Hausser, and J. Goudet. 2000. Microsatellites can be misleading: An empirical and simulation study. *Evolution* 54: 1414-1422.
Balloux, F., and N. Lugon-Moulin. 2002. The estimation of population differentiation with microsatellite markers. *Molecular Ecology* 11: 155-165.
Bandelt, H. J., P. Forster, and A. Rohl. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* 16: 37-48.
Banks, S. C., M. P. Piggott, J. E. Williamson, U. Bove, N. J. Holbrook, and L. B. Beheregaray. 2007. Oceanic variability and coastal topography shape genetic structure in a long-dispersing sea urchin. *Ecology* 88: 3055-3064.
Baranski, M., M. Rourke, S. Loughnan, C. Austin, and N. Robinson. 2006. Isolation and characterization of 125 microsatellite DNA markers in the blacklip abalone, *Haliotis rubra*. *Molecular Ecology Notes* 6: 740-746.
Barbará, T., C. Palma-Silva, G. M. Paggi, F. Bered, M. F. Fay, and C. Lexer. 2007. Cross-species transfer of nuclear microsatellite markers: potential and limitations. *Molecular Ecology* 16: 3759-3767.
Barber, P. H., S. R. Palumbi, M. V. Erdmann, and M. K. Moosa. 2000. Biogeography - A marine Wallace's line? *Nature* 406: 692-693.
Barber, P. H., S. R. Palumbi, M. V. Erdmann, and M. K. Moosa. 2002. Sharp genetic breaks among populations of *Haptosquilla pulchella* (Stomatopoda) indicate limits to larval transport: patterns, causes, and consequences. *Molecular Ecology* 11: 659-674.
Barnes, E. J. 1985. Eastern Cook Strait Region Circulation Inferred from Satellite-Derived, Sea-Surface, Temperature Data. *New Zealand Journal of Marine and Freshwater Research* 19: 405-411.
Becher, S. A., and R. Griffiths. 1997. Isolation and characterization of six polymorphic microsatellite loci in the European hedgehog Erinaceus europaeus. *Molecular Ecology* 6: 89-90.
Benzie, J. A. H. 1999. Genetic structure of coral reef organisms: Ghosts of dispersal past. *American Zoologist* 39: 131-145.
Benzie, J. A. H., E. Ballment, A. T. Forbes, N. T. Demetriades, K. Sugama, Haryanti, and S. Moria. 2002. Mitochondrial DNA variation in Indo-Pacific populations of the giant tiger prawn, *Penaeus monodon*. *Molecular Ecology* 11: 2553-2569.
Benzie, J. A. H., and S. T. Williams. 1996. Limitations in the genetic variation of hatchery produced batches of the giant clam, *Tridacna gigas*. *Aquaculture* 139: 225-241.
Bester, A. E., R. Slabbert, and M. E. D'Amato. 2004. Isolation and characterization of microsatellite markers in the South African abalone (*Haliotis midae*). *Molecular Ecology Notes* 4: 618-619.
Biermann, C. H. 1998. The molecular evolution of sperm bindin in six species of sea urchins (Echinoida : Strongylocentrotidae). *Molecular Biology and Evolution* 15: 1761-1771.
Bierne, N., F. Bonhomme, and P. David. 2003. Habitat preference and the marine-speciation paradox. *Proceedings of the Royal Society B-Biological Sciences* 270: 1399-1406.
Bird, C. E., B. S. Holland, B. W. Bowen, and R. J. Toonen. 2007. Contrasting phylogeography in three endemic Hawaiian limpets (*Cellana* spp.) with similar life histories. *Molecular Ecology* 16: 3173-3186.

Birky, C. W. 1995. Uniparental Inheritance of Mitochondrial and Chloroplast Genes - Mechanisms and Evolution. *Proceedings of the National Academy of Sciences of the United States of America* 92: 11331-11338.

Birky, C. W. 2001. The inheritance of genes in mitochondria and chloroplasts: Laws, mechanisms, and models. *Annual Review of Genetics* 35: 125-148.

Björklund, M. 2005. A method for adjusting allele frequencies in the case of microsatellite allele drop-out. *Molecular Ecology Notes* 5: 676-679.

Blake, S. G., N. J. Blake, M. J. Oesterling, and J. E. Graves. 1997. Genetic divergence and loss of diversity in two cultured populations of the bay scallop, *Argopecten irradians* (Lamarck, 1819). *Journal of Shellfish Research* 16: 55-58.

Bleiweiss, R. 1998. Slow rate of molecular evolution in high-elevation hummingbirds. *Proceedings of the National Academy of Sciences of the United States of America* 95: 612-616.

Boebel, O., T. Rossby, J. Lutjeharms, W. Zenk, and C. Barron. 2003. Path and variability of the Agulhas Return Current. *Deep-Sea Research Part II: Topical Studies in Oceanography* 50: 35-56.

Bohonak, A. J. 1999. Dispersal, gene flow, and population structure. *Quarterly Review of Biology* 74: 21-45.

Bonin, A., E. Bellemain, P. B. Eidesen, F. Pompanon, C. Brochmann, and P. Taberlet. 2004. How to track and assess genotyping errors in population genetics studies. *Molecular Ecology* 13: 3261-3273.

Borden, W. C., and C. A. Stepien. 2006. Discordant population genetic structuring of smallmouth bass, *Micropterus dolomieu* Lacèpede, in Lake Erie based on mitochondrial DNA sequences and nuclear DNA microsatellites. *Journal of Great Lakes Research* 32: 242-257.

Bowman, M. J., A. C. Kibblewhite, R. A. Murtagh, S. M. Chiswell, and B. G. Sanderson. 1983. Circulation and Mixing in Greater Cook Strait, New-Zealand. *Oceanologica Acta* 6: 383-391.

Bradford, J. M., P. P. Lapennas, R. A. Murtagh, F. H. Chang, and V. Wilkinson. 1986. Factors Controlling Summer Phytoplankton Production in Greater Cook Strait, New-Zealand. *New Zealand Journal of Marine and Freshwater Research* 20: 253-279.

Bradford-Grieve, J. M., R. C. Murdoch, and B. E. Chapman. 1993. Composition of Macrozooplankton Assemblages Associated with the Formation and Decay of Pulses within an Upwelling Plume in Greater Cook Strait, New-Zealand. *New Zealand Journal of Marine and Freshwater Research* 27: 1-22.

Braverman, J. M., R. R. Hudson, N. L. Kaplan, C. H. Langley, and W. Stephan. 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics* 140: 783-796.

Breen, P. A., and B. E. Adkins. 1980. Spawning in a British-Columbia Population of Northern Abalone, *Haliotis kamtschatkana*. *Veliger* 23: 177-179.

Bromham, L., and D. Penny. 2003. The modern molecular clock. *Nature Reviews Genetics* 4: 216-224.

Brookfield, J. F. Y. 1996. A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology* 5: 453-455.

Brown, L. D. 1991. Genetic variation and population structure in the blacklip abalone, *Haliotis rubra*. *Australian Journal of Marine & Freshwater Research* 42: 77-90.

Brown, L. D. 1995. Genetic evidence for hybridization between *Haliotis rubra* and *H. laevigata*. *Marine Biology* 123: 89-93.

Brown, L. D., and N. D. Murray. 1992a. Genetic relationships within the genus *Haliotis*. Pp. 19-23 *in* S. A. Shepherd, M. J. Tegner, and S. A. Guzman Del Proo, eds. *Abalone of the World: Biology, Fisheries and Culture*. Fishing News Books, Oxford.

Brown, L. D., and N. D. Murray. 1992b. Population genetics, gene flow and stock structure in *Haliotis rubra* and *Haliotis laevigata*. Pp. 24-33 *in* S. A. Shepherd, M. J. Tegner, and S. A.

G. del Proo, eds. *Abalone of the World: Biology, Fisheries and Culture*. Fishing New Books, London.
Brown, W. M., M. George, and A. C. Wilson. 1979. Rapid Evolution of Animal Mitochondrial-DNA. *Proceedings of the National Academy of Sciences of the United States of America* 76: 1967-1971.
Bruvo, R., N. K. Michiels, T. G. D'Souza, and H. Schulenburg. 2004. A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. *Molecular Ecology* 13: 2101-2106.
Burton, R. S., and M. J. Tegner. 2000. Enhancement of red abalone *Haliotis rufescens* stocks at San Miguel Island: Reassessing a success story. *Marine Ecology Progress Series* 202: 303-308.
Caizergues, A., O. Ratti, P. Helle, L. Rotelli, L. Ellison, and J. Y. Rasplus. 2003. Population genetic structure of male black grouse (*Tetrao tetrix* L.) in fragmented vs. continuous landscapes. *Molecular Ecology* 12: 2297-2305.
Campbell, A., J. Lessard, and G. S. Jamieson. 2003. Fecundity and seasonal reproduction of northern abalone, *Haliotis kamtschatkana*, in Barkley Sound, Canada. *Journal of Shellfish Research* 22: 811-818.
Capinpin, E. C., V. C. Encena, and N. C. Bayona. 1998. Studies on the reproductive biology of the Donkey's ear abalone, *Haliotis asinina* Linne. *Aquaculture* 166: 141-150.
Cassens, I., P. Mardulyn, and M. C. Milinkovitch. 2005. Evaluating intraspecific "Network" construction methods using simulated sequence data: Do existing algorithms outperform the global maximum parsimony approach? *Systematic Biology* 54: 363-372.
Cassens, I., K. Van Waerebeek, P. B. Best, E. A. Crespo, J. Reyes, and M. C. Milinkovitch. 2003. The phylogeography of dusky dolphins (*Lagenorhynchus obscurus*): a critical examination of network methods and rooting procedures. *Molecular Ecology* 12: 1781-1792.
Cassista, M. C., and M. W. Hart. 2007. Spatial and temporal genetic homogeneity in the Arctic surfclam (*Mactromeris polynyma*). *Marine Biology* 152: 569-579.
Castilla, J. C., and R. Guinez. 2000. Disjoint geographical distribution of intertidal and nearshore benthic invertebrates in the Southern Hemisphere. *Revista Chilena De Historia Natural* 73: 585-603.
Chambers, M. D., G. R. VanBlaricom, L. Hauser, F. Utter, and C. S. Friedman. 2006. Genetic structure of black abalone (*Haliotis cracherodii*) populations in the California islands and central California coast: Impacts of larval dispersal and decimation from withering syndrome. *Journal of Experimental Marine Biology and Ecology* 331: 173-185.
Chamary, J. V., J. L. Parmley, and L. D. Hurst. 2006. Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nature Review Genetics* 7: 98-108.
Chen, C., E. Durand, F. Forbes, and O. François. 2007. Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes* 7: 747-756.
Chiswell, S. M. 2000. The Wairarapa Coastal Current. *New Zealand Journal of Marine and Freshwater Research* 34: 303-315.
Chiswell, S. M. 2003. Circulation within the Wairarapa Eddy, New Zealand. *New Zealand Journal of Marine and Freshwater Research* 37: 691-704.
Chiswell, S. M. 2005. Mean and variability in the wairarapa and Hikurangi Eddies, New Zealand. *New Zealand Journal of Marine and Freshwater Research* 39: 121-134.
Chiswell, S. M., and J. D. Booth. 1999. Rock lobster *Jasus edwardsii* larval retention by the Wairarapa Eddy off New Zealand. *Marine Ecology-Progress Series* 183: 227-240.
Chiswell, S. M., and J. D. Booth. 2008. Sources and sinks of larval settlement in *Jasus edwardsii* around New Zealand: Where do larvae come from and where do they go? *Marine Ecology Progress Series* 354: 201-217.
Chiswell, S. M., and D. Roemmich. 1998. The East Cape Current and two eddies: a mechanism for larval retention? *New Zealand Journal of Marine and Freshwater Research* 32: 385-397.

Chiswell, S. M., and D. R. Schiel. 2001. Influence of along-shore advection and upwelling on coastal temperature at Kaikoura Peninsula, New Zealand. *New Zealand Journal of Marine and Freshwater Research* 35: 307-317.

Clark, A. G. 1990. Inference of haplotypes from PCR amplified samples of diploid populations. *Molecular Biology and Evolution* 7: 111-122.

Clark, N. L., G. D. Findlay, X. H. Yi, M. J. MacCoss, and W. J. Swanson. 2007. Duplication and selection on abalone sperm lysin in an allopatric population. *Molecular Biology and Evolution* 24: 2081-2090.

Clarke, C. B. 2001. Growth and survival of *Haliotis iris* in northern New Zealand, University of Auckland, Auckland, New Zealand.

Clavier, J. 1992. Fecundity and optimal sperm density for fertilization in ormer (*Haliotis tuberculata* L.) *in* S. A. Shepherd, M. J. Tegner, and S. A. Guzman Del Proo, eds. *Abalone of the World: Biology, Fisheries and Culture*. Fishing News Books, Oxford.

Clement, M., D. Posada, and K. A. Crandall. 2000. TCS: A computer program to estimate gene genealogies. *Molecular Ecology* 9: 1657-1659.

Coleman, A. W., and V. D. Vacquier. 2002. Exploring the phylogenetic utility of ITS sequences for animals: A test case for abalone (Haliotis). *Journal of Molecular Evolution* 54: 246-257.

Conod, N., J. P. Bartlett, B. S. Evans, and N. G. Elliott. 2002. Comparison of mitochondrial and nuclear DNA analyses of population structure in the blacklip abalone *Haliotis rubra* Leach. *Marine and Freshwater Research* 53: 711-718.

Counihan, R. T., D. C. McNamara, D. C. Souter, E. J. Jebreen, N. P. Preston, C. R. Johnson, and B. M. Degnan. 2001. Pattern, synchrony and predictability of spawning of the tropical abalone *Haliotis asinina* from Heron Reef, Australia. *Marine Ecology Progress Series* 213: 193-202.

Coustau, C., F. Renaud, C. Maillard, N. Pasteur, and B. Delay. 1991. Differential susceptibility to a trematode parasite among genotypes of the *Mytilus edulis*/*galloprovincialis* complex. *Genetical Research* 57: 207-212.

Crandall, E. D., M. A. Frey, R. K. Grosberg, and P. H. Barber. 2008. Contrasting demographic history and phylogeographical patterns in two Indo-Pacific gastropods. *Molecular Ecology* 17: 611-626.

Crandall, K. A., and A. R. Templeton. 1993. Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics* 134: 959-969.

Crow, J. F., and M. Kimura. 1970. *An Introduction to Population Genetics Theory* New York, USA.

Cruz, P., A. M. Ibarra, G. Fiore-Amaral, C. E. Galindo-Sa?nchez, and G. Mendoza-Carrio?n. 2005. Isolation of microsatellite loci in green abalone (*Haliotis fulgens*) and cross-species amplification in two other North American red (*Haliotis rufescens*) and pink (*Haliotis corrugata*) abalones. *Molecular Ecology Notes* 5: 857-859.

Daugherty, C. H., G. W. Gibbs, and R. A. Hitchmough. 1993. Mega-island or micro-continent? New Zealand and its fauna. *Trends in Ecology & Evolution* 8: 437-442.

Davis, A. R., and A. J. Butler. 1989. Direct observations of larval dispersal in the colonial ascidian *Podoclavella moluccensis* Sluiter - Evidence for closed populations. *Journal of Experimental Marine Biology and Ecology* 127: 189-203.

Davison, A., and S. Chiba. 2003. Laboratory temperature variation is a previously unrecognized source of genotyping error during capillary electrophoresis. *Molecular Ecology Notes* 3: 321-323.

Day, R. W., and A. E. Fleming. 1992. The determinants and measurement of abalone growth. Pp. 141-168 *in* S. A. Shepherd, M. J. Tegner, and S. A. G. del Proo, eds. *Abalone of the World: Biology, Fisheries and Culture*. Fishing News Books, Oxford.

Degnan, S. A., Imron, D. L. Geiger, and B. M. Degnan. 2006. Evolution in temperate and tropical seas: Disparate patterns in southern hemisphere abalone (Mollusca : Vetigastropoda : Haliotidae). *Molecular Phylogenetics and Evolution* 41: 249-256.

del Río Portilla, M. A. 2000. *Population genetics of the yellow abalone, Haliotis corrugata, in Cedros and San Benito Islands*. Pp. 508. *4th International Abalone Symposium*. Journal of Shellfish Research, Cape Town, South Africal.

del Río Portilla, M. A., and J. G. González-Avilés. 2001. Population genetics of the yellow abalone, *Haliotis corrugata*, in Cedros and San Benito Islands: A preliminary survey. *Journal of Shellfish Research* 20: 765-770.

Denver, D. R., K. Morris, M. Lynch, L. L. Vassilieva, and W. K. Thomas. 2000. High direct estimate of the mutation rate in the mitochondrial genome of Caenorhabditis elegans. *Science* 289: 2342-2344.

Desrosiers, R. R. 1993. A novel method to produce triploids in bivalve molluscs by the use of 6-dimethylaminopurine. *Journal of Experimental Marine Biology and Ecology* 170: 29-43.

DeWoody, J., J. D. Nason, and V. D. Hipkins. 2006. Mitigating scoring errors in microsatellite data from wild populations. *Molecular Ecology Notes* 6: 951-957.

Di Rienzo, A., A. C. Peterson, J. C. Garza, A. M. Valdes, M. Slatkin, and N. B. Freimer. 1994. Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences of the United States of America* 91: 3166-3170.

Díaz-Viloria, N., R. Pérez-Enríquez, G. Fiore-Amaral, R. S. Burton, and P. Cruz. 2008. Isolation and cross-amplification of microsatellites in pink abalone (*Haliotis corrugata*). *Molecular Ecology Resources* 8: 701-703.

Dixon, C. D., H. K. Gorfine, R. A. Officer, and M. Sporcic. 1998. Dispersal of tagged blacklip abalone, *Haliotis rubra*: Implications for stock assessment. *Journal of Shellfish Research* 17: 881-887.

Dixon, D. R. 1982. Aneuploidy in mussel embryos (*Mytilus edulis* L.) originating from a polluted dock. *Marine Biology Letters* 3: 155-161.

Dobretsov, S., and M. Wahl. 2001. Recruitment preferences of blue mussel spat (*Mytilus edulis*) for different substrata and microhabitats in the White Sea (Russia). *Hydrobiologia* 445: 27-35.

Dobzhansky, T. 1940. Speciation as a stage in evolutionary divergence. *American Naturalist* 74: 312-321.

Dollimore, J. M. 1977. The esterases of *Haliotis iris*. MSc, University of Auckland, Auckland, New Zealand.

Duda, T. F., and S. R. Palumbi. 1999. Population structure of the black tiger prawn, *Penaeus monodon*, among western Indian Ocean and western Pacific populations. *Marine Biology* 134: 705-710.

Dupanloup, I., S. Schneider, and L. Excoffier. 2002. A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology* 11: 2571-2581.

Durham, J. W. 1979. California's Cretaceous *Haliotis*. *Veliger* 21: 369-372.

Edmands, S., and D. C. Potts. 1997. Population genetic structure in brooding sea anemones (*Epiactis* spp.) with contrasting reproductive modes. *Marine Biology* 127: 485-498.

Edwards, S. V., and P. Beerli. 2000. Perspective: Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* 54: 1839-1854.

Ellegren, H. 2004. Microsatellites: Simple sequences with complex evolution. *Nature Reviews Genetics* 5: 435-445.

Estabrook, G. F., G. R. Smith, and T. E. Dowling. 2007. Body mass and temperature influence rates of mitochondrial DNA evolution in North American cyprinid fish. *Evolution* 61: 1176-1187.

Estes, J. A., D. R. Lindberg, and C. Wray. 2005. Evolution of large body size in abalones (*Haliotis*): patterns and implications. *Paleobiology* 31: 591-606.

Estoup, A., P. Jarne, and J. M. Cornuet. 2002. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology* 11: 1591-1604.

Evans, B., J. Bartlett, N. Sweijd, P. Cook, and N. G. Elliott. 2004a. Loss of genetic variation at microsatellite loci in hatchery produced abalone in Australia (*Haliotis rubra*) and South Africa (*Haliotis midae*). *Aquaculture* 233: 109-127.

Evans, B., N. Conod, and N. G. Elliott. 2001. Evaluation of microsatellite primer conservation in abalone. *Journal of Shellfish Research* 20: 1065-1070.
Evans, B., R. W. G. White, and N. G. Elliott. 2000. Characterization of microsatellite loci in the Australian Blacklip abalone (*Haliotis rubra*, Leach). *Molecular Ecology* 9: 1183-1184.
Evans, B. S., N. A. Sweijd, R. C. K. Bowie, P. A. Cook, and N. G. Elliott. 2004b. Population genetic structure of the perlemoen *Haliotis midae* in South Africa: Evidence of range expansion and founder events. *Marine Ecology Progress Series* 270: 163-172.
Excoffier, L. 2003. Analysis of Population Subdivision. Pp. 713-750 *in* D. J. Balding, M. Bishop, and C. Cannings, eds. *Handbook of Statistical Genetics*. John Wiley & Sons, Ltd., Chichester (UK).
Excoffier, L. 2004. Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. *Molecular Ecology* 13: 853-864.
Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1: 47-50.
Excoffier, L., and P. E. Smouse. 1994. Using allele frequencies and geographic subdivision to reconstruct gene trees within a species: Molecular Variance Parsimony. *Genetics* 136: 343-359.
Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. 131: 479-491.
FAO. 2000. FISHSTAT Plus: Universal software for fishery statistical time series. FAO Fisheries Department, Fishery Information, Data and Statistcs Unit.
Farrell, T. M., D. Bracher, and J. Roughgarden. 1991. Cross-Shelf transport causes recruitment to intertidal populations in central California. *Limnology and Oceanography* 36: 279-288.
Felsenstein, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Moleular Evolution* 17: 368-376.
Felsenstein, J. 2004. *Inferring Phylogenies*. Sinauer Associates, Inc., Sunderland, MA.
Féral, J. P. 2002. How useful are the genetic markers in attempts to understand and manage marine biodiversity? *Journal of Experimental Marine Biology and Ecology* 268: 121-145.
Fielding, P. J. 1995. A preliminary investigation of abalone *Haliotis midae* resources along the Transkei coast, South Africa. *South African Journal of Marine Science-Suid-Afrikaanse Tydskrif Vir Seewetenskap* 15: 253-261.
Fleming, C. A. 1979. *The Geological History of New Zealand*. Auckland University Press, Auckland.
Flot, J. F. 2007. CHAMPURU 1.0: A computer software for unraveling mixtures of two DNA sequences of unequal lengths. *Molecular Ecology Notes* 7: 974-977.
Flot, J. F., A. Tillier, S. Samadi, and S. Tillier. 2006. Phase determination from direct sequencing of length-variable DNA regions. *Molecular Ecology Notes* 6: 627-630.
Francis, M. P. 1996. Geographic distribution of marine reef fishes in the New Zealand region. 30: 35-55.
Frank, S. A. 2000. Sperm competition and female avoidance of polyspermy mediated by sperm-egg biochemistry. *Evolutionary Ecology Research* 2: 613-625.
Fridberger, A., J. Sundelin, V. D. Vacquier, and P. A. Peterson. 1985. Amino-acid sequence of an egg-lysin protein from abalone spermatozoa that solubilizes the vitelline layer. *Journal of Biological Chemistry* 260: 9092-9099.
Frusin, A. 1982. Electrophoretic study of some paua (*Haliotis iris)* proteins, Victoria University, Wellington, New Zealand.
Fu, Y. X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147: 915-925.
Fu, Y. X., and W. H. Li. 1993. Statistical tests of neutrality of mutations. *Genetics* 133: 693-709.

Gaffney, P. M. 2000. Molecular tools for understanding population structure in Antarctic species. *Antarctic Science* 12: 288-296.
Gaffney, P. M., V. P. Rubin, D. Hedgecock, D. A. Powers, G. Morris, and L. Hereford. 1996. Genetic effects of artificial propagation: Signals from wild and hatchery populations of red abalone in California. *Aquaculture* 143: 257-266.
Gaggiotti, O. E., O. Lange, K. Rassmann, and C. Gliddon. 1999. A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. *Molecular Ecology* 8: 1513-1520.
Gaines, S. D., and M. D. Bertness. 1992. Dispersal of juveniles and variable recruitment in sessile marine species. *Nature* 360: 579-580.
Galindo, B. E., G. W. Moy, W. J. Swanson, and V. D. Vacquier. 2002. Full-length sequence of VERL, the egg vitelline envelope receptor for abalone sperm lysin. *Gene* 288: 111-117.
Galindo, B. E., V. D. Vacquier, and W. J. Swanson. 2003. Positive selection in the egg receptor for abalone sperm lysin. *Proceedings of the National Academy of Sciences of the United States of America* 100: 4639-4643.
Gallardo-Escárate, C., J. Álvarez-Borrego, M. A. del Río Portilla, and V. Kober. 2004. Karyotype of pacific red abalone *Haliotis rufescens* (Archaeogastropoda: Haliotidae) using image analysis. *Journal of Shellfish Research* 23: 205-209.
Galtier, N., F. Depaulis, and N. H. Barton. 2000. Detecting bottlenecks and selective sweeps from DNA sequence polymorphism. *Genetics* 155: 981-987.
Gardner, J. P. A., A. Pande, R. F. Eyles, and R. G. Wear. 1996. Biochemical genetic variation among populations of the greenshell mussel, Perna canaliculus, from New Zealand: Preliminary findings. *Biochemical Systematics and Ecology* 24: 763-774.
Gardner, J. P. A., and D. O. F. Skibinski. 1991. Biological and physical factors influencing genotype-dependent mortality in hybrid mussel populations *Marine Ecology Progress Series* 71: 235-243.
Garrick, R. C., R. J. Dyer, L. B. Beheregaray, and P. Sunnucks. 2008. Babies and bathwater: a comment on the premature obituary for nested clade phylogeographical analysis. *Molecular Ecology* 17: 1401-1403.
Gavrilets, S., and D. Waxman. 2002. Sympatric speciation by sexual conflict. *Proceedings of the National Academy of Sciences of the United States of America* 99: 10533-10538.
Geiger, D. L. 1999. A total evidence cladistic analysis of the Haliotidae (Gastropoda: Vetigastropoda). Doctor of Philosophy, University of Southern California, Los Angeles.
Geiger, D. L. 2000. Distribution and biogeography of the recent Haliotidae (Gastropoda: Vetigastropoda) world-wide. *Bollettino Malacologico, Roma* 35: 57-120.
Geiger, D. L., and L. T. Groves. 1999. Review of fossil abalone (Gastropoda : Vetigastropoda : Haliotidae) with comparison to Recent species. *Journal of Paleontology* 73: 872-885.
Geiger, D. L., and C. E. Thacker. 2005. Molecular phylogeny of Vetigastropoda reveals non-monophyletic Scissurellidae, Trochoidea, and Fissurelloidea. *Molluscan Research* 25: 47-55.
Gemmell, N. J., and S. Akiyama. 1996. An efficient method for the extraction of DNA from vertebrate tissues. *Trends in Genetics* 12: 338-339.
Genin, A., J. S. Jaffe, R. Reef, C. Richter, and P. J. S. Franks. 2005. Swimming against the flow: A mechanism of zooplankton aggregation. *Science* 308: 860-862.
Geyer, L. B., and S. R. Palumbi. 2003. Reproductive character displacement and the genetics of gamete recognition in tropical sea urchins. *Evolution* 57: 1049-1060.
Gilg, M. R., and T. J. Hilbish. 2000. The relationship between allele frequency and tidal height in a mussel hybrid zone: a test of the differential settlement hypothesis. *Marine Biology* 137: 371-378.
Gilg, M. R., and T. J. Hilbish. 2003. The geography of marine larval dispersal: Coupling genetics with fine-scale physical oceanography. *Ecology* 84: 2989-2998.

Gillooly, J. F., A. P. Allen, G. B. West, and J. H. Brown. 2005. The rate of DNA evolution: Effects of body size and temperature on the molecular clock. *Proceedings of the National Academy of Sciences of the United States of America* 102: 140-145.

Gillooly, J. F., M. W. McCoy, and A. P. Allen. 2007. Effects of metabolic rate on protein evolution. *Biology Letters* 3: 655-659.

Gluyas-Millan, M. G., and J. Talavera-Maya. 2003. Size and age composition of the populations of abalone *Haliotis fulgens* and *H. corrugata* of Bahia Tortugas, Baja California Sur, Mexico. *Ciencias Marinas* 29: 89-101.

Goldson, A. J., R. N. Hughes, and C. J. Gliddon. 2001. Population genetic consequences of larval dispersal mode and hydrography: a case study with bryozoans. *Marine Biology* 138: 1037-1042.

Goldstien, S. J. 2005. Phylogeography of the *Cellana* limpets of New Zealand: Investigating barriers to marine dispersal and historical biogeography. Ph. D., University of Canterbury, Christchurch.

Goldstien, S. J., N. J. Gemmell, and D. R. Schiel. 2006a. Molecular phylogenetics and biogeography of the nacellid limpets of New Zealand (Mollusca : Patellogastropoda). *Molecular Phylogenetics and Evolution* 38: 261-265.

Goldstien, S. J., D. R. Schiel, and N. J. Gemmell. 2006b. Comparative phylogeography of coastal limpets across a marine disjunction in New Zealand. *Molecular Ecology* 15: 3259-3268.

Goudet, J. 2002. Fstat 2.9.3.2. Lausanne University.

Grantham, B. A., G. L. Eckert, and A. L. Shanks. 2003. Dispersal potential of marine invertebrates in diverse habitats. *Ecological Applications* 13: S108-S116.

Graur, D., and W. Li. 2000. *Fundamentals of molecular evolution*. Sinauer Associates, Inc., Sunderland, MA, USA.

Gruenthal, K. M., L. K. Acheson, and R. S. Burton. 2007. Genetic structure of natural populations of California red abalone (*Haliotis rufescens*) using multiple genetic markers. *Marine Biology* 152: 1237-1248.

Gruenthal, K. M., and R. S. Burton. 2005. Genetic diversity and species identification in the endangered white abalone (*Haliotis sorenseni*). *Conservation Genetics* 6: 929-939.

Gruenthal, K. M., and R. S. Burton. 2008. Genetic structure of natural populations of the California black abalone (*Haliotis cracherodii* Leach, 1814), a candidate for endangered species status. *Journal of Experimental Marine Biology and Ecology* 355: 47-58.

Guo, S. W., and E. A. Thompson. 1992. Performing the exact test of hardy-weinberg proportion for multiple alleles. *Biometrics* 48: 361-372.

Gutiérrez-Gonzalez, J. L. 2000. *Genetic variability of the blue abalone Haliotis fulgens in the west coast of Baja, California*. Pp. 517. *4th International Abalone Symposium*. Journal of Shellfish Research, Cape Town, South Africa.

Gutiérrez-Gonzalez, J. L., P. Cruz, M. A. del Rio-Portilla, and R. Perez-Enriquez. 2007. Genetic structure of green abalone *Haliotis fulgens* population off Baja California, Mexico. *Journal of Shellfish Research* 26: 839-846.

Gutiérrez-Gonzalez, J. L., and R. Perez-Enriquez. 2005. A genetic evaluation of stock enhancement of blue abalone *Haliotis fulgens* in Baja California, Mexico. *Aquaculture* 247: 233-242.

Gutow, L., J. Strahl, C. Wiencke, H. D. Franke, and R. Saborowski. 2006. Behavioural and metabolic adaptations of marine isopods to the rafting life style. *Marine Biology* 149: 821-828.

Hamm, D. E., and R. S. Burton. 2000. Population genetics of black abalone, *Haliotis cracherodii*, along the central California coast. *Journal of Experimental Marine Biology and Ecology* 254: 235-247.

Hancock, B. 2000. Genetic subdivision of Roe's abalone, *Haliotis roei* Grey (Mollusca: Gastropoda), in south-western Australia. *Marine and Freshwater Research* 51: 679-687.

Haszprunar, G. 1988. On the origin and evolution of major gastropod groups, with special reference to the Streptoneura. *Journal of Molluscan Studies* 54.
Hara, M., and M. Sekino. 2005. Genetic difference between Ezo-awabi *Haliotis discus hannai* and Kuro-awabi *H. discus discus* populations: Microsatellite-based population analysis in Japanese abalone. *Fisheries Science* 71: 754-766.
Hara, M., and M. Sekino. 2007. Parentage testing for hatchery-produced abalone *Haliotis discus hannai* based on microsatellite markers: preliminary evaluation of early growth of selected strains in mixed family farming. *Fisheries Science* 73: 831-836.
Hare, M. P., and S. R. Palumbi. 1999. The accuracy of heterozygous base calling from diploid sequence and resolution of haplotypes using allele-specific sequencing. *Molecular Ecology* 8: 1750-1752.
Harpending, H. C. 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology* 66: 591-600.
Harpending, H. C., S. T. Sherry, A. R. Rogers, and M. Stoneking. 1993. The Genetic structure of ancient human populations. *Current Anthropology* 34: 483-496.
Harris, T. F. W. 1990. *Greater Cook Strait: Form and Flow*. New Zealand Oceanographic Institute: DSIR Marine and Freshwater, Wellington.
Hartl, D. L., and A. G. Clark. 1997. *Principles of Population Genetics*. Sinauer Associates, Inc., Sunderland, MA USA.
Haszprunar, G. 1988. On the origin and evolution of major gastropod groups, with special reference to the Streptoneura. *Journal of Molluscan Studies* 54.
Hauser, L., G. J. Adcock, P. J. Smith, J. H. B. Ramirez, and G. R. Carvalho. 2002. Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (*Pagrus auratus*). *Proceedings of the National Academy of Sciences of the United States of America* 99: 11742-11747.
Haygood, R. 2004. Sexual conflict and protein polymorphism. *Evolution* 58: 1414-1423.
Heath, R. A. 1970. Hydrology and circulation in central and southern Cook Strait, New Zealand. *New Zealand Journal of Marine and Freshwater Research* 5: 178-199.
Heath, R. A. 1972a. Oceanic upwelling produced by northerly winds on the north Canterbury coast, New Zealand. *New Zealand Journal of Marine and Freshwater Research* 6: 343-351.
Heath, R. A. 1972b. Wind-derived water motions off the East Coast of New Zealand. *New Zealand Journal of Marine and Freshwater Research* 6: 352-364.
Heath, R. A. 1978. Semidiurnal tides in Cook Strait. *New Zealand Journal of Marine and Freshwater Research* 12: 87-97.
Heath, R. A. 1985. A review of the physical oceanography of the seas around New Zealand – 1982. *New Zealand Journal of Marine and Freshwater Research* 19: 79-124.
Heath, R. A. 1986. In which direction is the mean flow through Cook Strait, New-Zealand? Evidence of 1 to 4 week variability. *New Zealand Journal of Marine and Freshwater Research* 20: 119-137.
Hedgecock, D. 1986. Is gene flow from pelagic larval dispersal important in the adaptation and evolution of marine invertebrates. *Bulletin of Marine Science* 39: 550-564.
Hedgecock, D. 1994. Does variance in reproductive success limit effective population size of marine organisms? . Pp. 122-134 *in* R. A. Beaumont, ed. *Genetics and evolution of aquatic organisms*. Chapman & Hall, London.
Hedgecock, D., G. Li, S. Hubert, K. Bucklin, and V. Ribes. 2004. Widespread null alleles and poor cross-species amplification of microsatellite DNA loci cloned from the Pacific oyster, *Crassostrea gigas*. *Journal of Shellfish Research* 23: 379-385.
Hedgecock, D., and F. Sly. 1990. Genetic drift and effective population sizes of hatchery-propagated stocks of the Pacific oyster, *Crassostrea gigas*. *Aquaculture* 88: 21-38.
Hedrick, P. 2005. Large variance in reproductive success and the N_e/N ratio. *Evolution* 59: 1596-1599.

Hedrick, P. W. 1999. Perspective: Highly variable loci and their interpretation in evolution and conservation. *Evolution* 53: 313-318.
Held, C. 2001. No evidence for slow-down of molecular substitution rates at subzero temperatures in Antarctic serolid isopods (Crustacea, Isopoda, Serolidae). *Polar Biology* 24: 497-501.
Hellberg, M. E. 2006. No variation and low synonymous substitution rates in coral mtDNA despite high nuclear variation. *Bmc Evolutionary Biology* 6: 24.
Hellberg, M. E. 1996. Dependence of gene flow on geographic distance in two solitary corals with different larval dispersal capabilities. *Evolution* 50: 1167-1175.
Hellberg, M. E., G. W. Moy, and V. D. Vacquier. 2000. Positive selection and propeptide repeats promote rapid interspecific divergence of a gastropod sperm protein. *Molecular Biology and Evolution* 17: 458-466.
Hellberg, M. E., and V. D. Vacquier. 1999. Rapid evolution of fertilization selectivity and lysin cDNA sequences in teguline gastropods. *Molecular Biology and Evolution* 16: 839-848.
Hernández-Ibarra, N. K., C. Márquez, J. L. Ramírez, and A. M. Ibarra. 2004. Comparative karyotypes of two northeastern pacific abalone species (*Haliotis fulgens* Philippi and *Haliotis rufescens* Swainson). *Journal of Shellfish Research* 23: 861-865.
Hertlein, L. G. 1937. *Haliotis koticki*, a new species from the lower Miocene of California. *Bulletin (Southern California Academy of Sciences)* 36: 93-97.
Hey, J., and C. A. Machado. 2003. The study of structured populations: New hope for a difficult and divided science. *Nature Reviews Genetics* 4: 535-543.
Highsmith, R. C. 1985. Floating and algal rafting as potential dispersal mechanisms in brooding invertebrates. *Marine Ecology Progress Series* 25: 169-179.
Hilbish, T. J., E. W. Carson, J. R. Plante, L. A. Weaver, and M. R. Gilg. 2002. Distribution of *Mytilus edulis*, *M. galloprovincialis*, and their hybrids in open-coast populations of mussels in southwestern England. *Marine Biology* 140: 137-142.
Hobday, A. J., M. J. Tegner, and P. L. Haaker. 2001. Over-exploitation of a broadcast spawning marine invertebrate: Decline of the white abalone. *Reviews in Fish Biology and Fisheries* 10: 493-514.
Hoffman, J. I., and W. Amos. 2005. Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Molecular Ecology* 14: 599-612.
Holmquist, J. G. 1994. Benthic macroalgae as a dispersal mechanism for fauna - influence of marine tumbleweed. *Journal of Experimental Marine Biology and Ecology* 180: 235-251.
Hooker, S. H., and R. G. Creese. 1995. Reproduction of Paua, *Haliotis iris* Gmelin 1791 (Mollusca, Gastropoda), in north eastern New Zealand. *Marine and Freshwater Research* 46: 617-622.
Hoshikawa, H., Y. Sakai, and A. Kijima. 1998. Growth characteristics of the hybrid between pinto abalone, *Haliotis kamtschatkana* Jonas, and exo abalone, *H. discus hannai* Ino, under high and low temperature. *Journal of Shellfish Research* 17: 673-677.
http://www.fish.govt.nz. 2006. Status of Fisheries: Paua.
Huang, B., Z. Chai, P. J. Hanna, and K. H. Gough. 1997. Molecular sequences of two minisatellites in blacklip abalone, *Haliotis rubra*. *Electrophoresis* 18: 1653-1659.
Huang, B., and P. J. Hanna. 1998. Identification of three polymorphic microsatellite loci in blacklip abalone, *Haliotis rubra* (Leach), and detection in other abalone species. *Journal of Shellfish Research* 17: 795-799.
Huang, B. X., R. Peakall, and P. J. Hanna. 2000. Analysis of genetic structure of blacklip abalone (*Haliotis rubra*) populations using RAPD, minisatellite and microsatellite markers. *Marine Biology* 136: 207-216.
Hudson, R. R. 1990. Gene genalogies and the coalescent process. Pp. 1-44 *in* D. Futuyma, and J. Antonovics, eds. *Oxford Surveys in Evolutionary Biology*. Oxford University Press, USA.
Hume, T. M., R. G. Bell, W. P. Delange, T. R. Healy, D. M. Hicks, and R. M. Kirk. 1992. Coastal oceanography and sedimentology in New Zealand, 1967-91. *New Zealand Journal of Marine and Freshwater Research* 26: 1-36.

Ibarra, A. M., N. K. Hernandez-Ibarra, P. Cruz, R. Perez-Enriquez, S. Avila, and J. L. Ramirez. 2005. Genetic certification of presumed hybrids of blue X red abalone (*Haliotis fulgens* Philippi and *H. rufescens* Swainson). *Aquaculture Research* 36: 1356-1368.

Imron, B. Jeffrey, P. Hale, B. M. Degnan, and S. M. Degnan. 2007. Pleistocene isolation and recent gene flow in *Haliotis asinina*, an Indo-Pacific vetigastropod with limited dispersal capacity. *Molecular Ecology* 16: 289-304.

Innan, H., and M. Nordborg. 2003. The extent of linkage disequilibrium and haplotype sharing around a polymorphic site. *Genetics* 165: 437-444.

Jablonski, D. 1986. Larval ecology and macroevolution in marine invertebrates. *Bulletin of Marine Science* 39: 565-587.

Jaeckle, W. B., and D. T. Manahan. 1992. Experimental Manipulations of the Organic Composition of Seawater: Implications for Studies of Energy Budgets in Marine Invertebrate Larvae. *Journal of Experimental Marine Biology and Ecology* 156: 273-284.

Jiang, L., W. L. Wu, and P. C. Huang. 1995. The mitochondrial DNA of taiwan abalone *Haliotis diversicolor* Reeve, 1846 (Gastropoda: Archaeogastropoda: Haliotidae). *Molecular Marine Biology and Biotechnology* 4: 353-364.

Jonasson, J., S. E. Stefansson, A. Gudnason, and A. Steinarsson. 1999. Genetic variation for survival and shell length of cultured red abalone (*Haliotis rufescens*) in Iceland. *Journal of Shellfish Research* 18: 621-625.

Kalinowski, S. 2002a. How many alleles per locus should be used to estimate genetic distances? *Heredity* 88: 62-65.

Kalinowski, S. T. 2002b. Evolutionary and statistical properties of three genetic distances. *Molecular Ecology* 11: 1263-1273.

Kalinowski, S. T. 2005. Do polymorphic loci require large sample sizes to estimate genetic distances? *Heredity* 94: 33-36.

Kamei, N., and C. G. Glabe. 2003. The species-specific egg receptor for sea urchin sperm adhesion is EBR1, a novel ADAMTS protein. *Genes & Development* 17: 2502-2507.

Kawecki, T. J., and D. Ebert. 2004. Conceptual issues in local adaptation. *Ecology Letters* 7: 1225-1241.

Kimura, M., and J. F. Crow. 1964. The number of alleles that can be maintained in a finite population. *Genetics* 49: 725-738.

Kimura, M., and T. Ohta. 1978. Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proceedings of the National Academy of Sciences of the United States of America* 75: 2868-2872.

Kingman, J. F. C. 1982. The coalescent. *Stochastic Processes and their Applications* 13.

Kingsford, M. J., J. M. Leis, A. Shanks, K. C. Lindeman, S. G. Morgan, and J. Pineda. 2002. Sensory environments, larval abilities and local self-recruitment. *Bulletin of Marine Science* 70: 309-340.

Kirby, V. L., R. Villa, and D. A. Powers. 1998. Identification of microsatellites in the California red abalone, *Haliotis rufescens*. *Journal of Shellfish Research* 17: 801-804.

Klinbunga, S., P. Pripue, N. Khamnamtong, N. Puanglarp, A. Tassanakajon, P. Jarayabhand, I. Hirono, T. Aoki, and P. Menasveta. 2003. Genetic diversity and molecular markers of the tropical abalone (*Haliotis asinina*) in Thailand. *Marine Biotechnology* 5: 505-517.

Kloda, J. M., P. D. G. Dean, C. Maddren, D. W. MacDonald, and S. Mayes. 2008. Using principle component analysis to compare genetic diversity across polyploidy levels within plant complexes: An example from British Restharrows (*Ononis spinosa* and *Ononis repens*). *Heredity* 100: 253-260.

Knowles, L. L., and W. P. Maddison. 2002. Statistical phylogeography. *Molecular Ecology* 11: 2623-2635.

Kondo, R., Y. Satta, E. T. Matsuura, H. Ishiwa, N. Takahata, and S. I. Chigusa. 1990. Incomplete maternal transmission of mitochondrial-DNA in drosophila. *Genetics* 126: 657-663.

Kong, L. F., and Q. Li. 2007. Genetic comparison of cultured and wild populations of the clam *Coelomactra antiquata* (Spengler) in China using AFLP markers. *Aquaculture* 271: 152-161.

Korpelainen, H. 2004. The evolutionary processes of mitochondrial and chloroplast genomes differ from those of nuclear genomes. *Naturwissenschaften* 91: 505-518.

Kosman, E., and K. J. Leonard. 2005. Similarity coefficients for molecular markers in studies of genetic relationships between individuals for haploid, diploid, and polyploid species. *Molecular Ecology* 14: 415-424.

Kraytsberg, Y., M. Schwartz, T. A. Brown, K. Ebralidse, W. S. Kunz, D. A. Clayton, J. Vissing, and K. Khrapko. 2004. Recombination of Human Mitochondrial DNA. *Science* 304: 981.

Kresge, N., V. D. Vacquier, and C. D. Stout. 2001. Abalone lysin: the dissolving and evolving sperm protein. *Bioessays* 23: 95-103.

Kube, P. D., S. A. Appleyard, and N. G. Elliott. 2007. Selective breeding greenlip abalone (*Haliotis laevigata*): Preliminary results and issues. *Journal of Shellfish Research* 26: 821-824.

Laing, A., and S. Chiswell. 2003. The Ocean Medium. Pp. 24 - 31 *in* N. Andrew, and M. Francis, eds. *The Living Reef: The Ecology of New Zealand's Rocky Reefs*. Craig Potton Publishing, Nelson, NZ.

Lander, E. S., L. M. Linton, B. Birren, et al.. Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.

Lanfear, R., J. A. Thomas, J. J. Welch, T. Brey, and L. Bromham. 2007. Metabolic rate does not calibrate the molecular clock. *Proceedings of the National Academy of Sciences of the United States of America* 104: 15388-15393.

Largier, J. L., P. Chapman, W. T. Peterson, and V. P. Swart. 1992. The western Agulhas Bank: circulation, stratification and ecology. *South African Journal of Marine Science* 12: 319-339.

Lee, H. J., and E. G. Boulding. 2007. Mitochondrial DNA variation in space and time in the northeastern Pacific gastropod, *Littorina keenae*. *Molecular Ecology* 16: 3084-3103.

Lee, Y., and V. D. Vacquier. 1992. The divergence of species-specific abalone sperm lysins Is promoted by positive darwinian selection. *Biological Bulletin* 182: 97-104.

Lee, Y. H., T. Ota, and V. D. Vacquier. 1995. Positive selection is a general phenomenon in the evolution of abalone sperm lysin. *Molecular Biology and Evolution* 12: 231-238.

Lee, Y. H., and V. D. Vacquier. 1995. Evolution and systematics in Haliotidae (Mollusca: Gastropoda): Inferences from DNA sequences of sperm lysin. *Marine Biology* 124: 267-278.

Leiva, G. E., and J. C. Castilla. 2001. A review of the world marine gastropod fishery: Evolution of catches, management and the Chilean experience. *Reviews in Fish Biology and Fisheries* 11: 283-300.

Lemaire, C., J. J. Versini, and F. Bonhomme. 2005. Maintenance of genetic differentiation across a transition zone in the sea: discordance between nuclear and cytoplasmic markers. *Journal of Evolutionary Biology* 18: 70-80.

Lenormand, T. 2002. Gene flow and the limits to natural selection. *Trends in Ecology and Evolution* 17: 183-189.

Levin, L. A. 2006. Recent progress in understanding larval dispersal: new directions and digressions. *Integrative and Comparative Biology* 46: 282-297.

Levitan, D. R. 2004. Density-dependent sexual selection in external fertilizers: Variances in male and female fertilization success along the continuum from sperm limitation to sexual conflict in the sea urchin Strongylocentrotus franciscanus. *American Naturalist* 164: 298-309.

Levitan, D. R., and D. L. Ferrell. 2006. Selection on gamete recognition proteins depends on sex, density, and genotype frequency. *Science* 312: 267-269.

Li, N., and M. Stephens. 2003. Modelling linkage disequilibrium, and identifying recombination hotsponts using snp data. *Genetics* 165: 2213-2233.
Li, Q., C. Park, T. Endo, and A. Kijima. 2004. Loss of genetic variation at microsatellite loci in hatchery strains of the Pacific abalone (*Haliotis discus hannai*). *Aquaculture* 235: 207-222.
Li, Q., C. Park, and A. Kijima. 2002a. Isolation and characterization of microsatellite loci in the Pacific abalone, *Haliotis discus hannai*. *Journal of Shellfish Research* 21: 811-815.
Li, Q., J. Shu, R. H. Yu, and C. Y. Tian. 2007a. Genetic variability of cultured populations of the Pacific abalone (*Haliotis discus hannai* Ino) in China based on microsatellites. *Aquaculture Research* 38: 981-990.
Li, Y., X. Li, and J. G. Qin. 2007b. Triploidy induction in Australian greenlip abalone *Haliotis laevigata* (Donovan) with cytochalasin B. *Aquaculture Research* 38: 487-492.
Li, Y.-C., A. B. Korol, T. Fahima, A. Beiles, and E. Nevo. 2002b. Microsatellites: Genomic distribution, putative functions and mutational mechanisms: A review. *Molecular Ecology* 11: 2453-2465.
Lindberg, D. R. 1992. Evolution, distribution and systematics of Haliotidae. Pp. 3 - 18 *in* S. A. Shepherd, M. J. Tegner, and S. A. Guzman Del Proo, eds. *Abalone of the World: Biology, Fisheries and Culture*. Fishing News Books, London.
Liu, W., M. Heasman, and R. Simpson. 2004. Induction and evaluation of triploidy in the Australian blacklip abalone, *Haliotis rubra*: A preliminary study. *Aquaculture* 233: 79-92.
Lodish, H., D. Baltimore, A. Berk, S. L. Zipursky, P. Matsudaira, and J. Darnell. 1995. *Molecular Cell Biology*. Scientific American Books, Inc., New York.
Lourie, S. A., D. M. Green, and A. C. J. Vincent. 2005. Dispersal, habitat differences, and comparative phylogeography of Southeast Asian seahorses (Syngnathidae : Hippocampus). *Molecular Ecology* 14: 1073-1094.
Lucas, T., M. Macbeth, S. M. Degnan, W. Knibb, and B. M. Degnan. 2006. Heritability estimates for growth in the tropical abalone *Haliotis asinina* using microsatellites to assign parentage. *Aquaculture* 259: 146-152.
Luttikhuizen, P. C., M. Stift, P. Kuperus, and P. H. Van Tienderen. 2007. Genetic diversity in diploid vs. tetraploid *Rorippa amphibia* (Brassicaceae). *Molecular Ecology* 16: 3544-3553.
Lyon, J. D., and V. D. Vacquier. 1999. Interspecies chimeric sperm lysins identify regions mediating species-specific recognition of the abalone egg vitelline envelope. *Developmental Biology* 214: 151-159.
Maldonado, R., A. M. Ibarra, J. L. Ramírez, S. Avila, J. E. Vázquez, and L. M. Badillo. 2001. Induction of triploidy in pacific red abalone (*Haliotis rufescens*). *Journal of Shellfish Research* 20: 1071-1075.
Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research* 27: 209-220.
Mariette, S., V. Le Corre, F. Austerlitz, and A. Kremer. 2002. Sampling within the genome for measuring within-population diversity: trade-offs between markers. *Molecular Ecology* 11: 1145-1156.
Marshall, J. L., M. L. Arnold, and D. J. Howard. 2002. Reinforcement: The road not taken. *Trends in Ecology and Evolution* 17: 558-563.
Martin, A. P., and S. R. Palumbi. 1993. Body size, metabolic-rate, generation time, and the molecular clock. *Proceedings of the National Academy of Sciences of the United States of America* 90: 4087-4091.
Maynard, B. T., L. J. Kerr, J. M. McKiernan, E. S. Jansen, and P. J. Hanna. 2005. Mitochondrial DNA sequence and gene organization in Australian backup abalone *Haliotis rubra* (Leach). *Marine Biotechnology* 7: 645-658.
McCormick, T. B., L. M. Buckley, J. Brogan, and L. M. Perry. 2008. Drift macroalgae as a potential dispersal mechanism for the white abalone *Haliotis sorenseni*. *Marine Ecology Progress Series* 362: 225-232.

McDonald, J. H., and M. Kreitman. 1991. Adaptive protein evolution at the Adh locus in Drosophila. *Nature* 351: 652-654.
McShane, P. E. 1992. Early life history of abalone: a review. Pp. 120-138 *in* S. A. Shepherd, M. J. Tegner, and S. A. Guzman Del Proo, eds. *Abalone of the World: Biology, Fisheries and Culture*. Fishing News Books 1992, Oxford.
McShane, P. E. 1996. Patch dynamics and effects of exploitation on abalone (*Haliotis iris*) populations. *Fisheries Research* 25: 191-199.
McShane, P. E., and J. R. Naylor. 1995a. Depth can affect postsettlement survival of Haliotis iris (Mollusca, Gastropoda). *Journal of Experimental Marine Biology and Ecology* 187: 1-12.
McShane, P. E., and J. R. Naylor. 1995b. Small-scale spatial variation in growth, size at maturity, and yield- and egg-per-recruit relations in the New Zealand abalone *Haliotis iris*. *New Zealand Journal of Marine and Freshwater Research* 29: 603-612.
McShane, P. E., and J. R. Naylor. 1997. *Direct estimation of natural mortality of the New Zealand abalone, Haliotis iris (Note)*. Pp. 135-137. *New Zealand Journal of Marine and Freshwater Research*.
McShane, P. E., D. R. Schiel, S. F. Mercer, and T. Murray. 1994. Morphometric variation in *Haliotis iris* (Mollusca, Gastropoda) - Analysis of 61 Populations. *New Zealand Journal of Marine and Freshwater Research* 28: 357-364.
McShane, P. E., and M. G. Smith. 1988. Measuring abundance of juvenile abalone, *Haliotis rubra* Leach (Gastropoda, Haliotidae); Comparison of a novel method with 2 other methods. *Australian Journal of Marine and Freshwater Research* 39: 331-336.
Metz, E. C., G. Gomez-Gutierrez, and V. D. Vacquier. 1998a. Mitochondrial DNA and bindin gene sequence evolution among allopatric species of the sea urchin genus *Arbacia*. *Molecular Biology and Evolution* 15: 185-195.
Metz, E. C., and S. R. Palumbi. 1996. Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. *Molecular Biology and Evolution* 13: 397-406.
Metz, E. C., R. Robles-Sikisaka, and V. D. Vacquier. 1998b. Nonsynonymous substitution in abalone sperm fertilization genes exceeds substitution in introns and mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America* 95: 10676-10681.
Miller, K. J., C. N. Mundy, and W. L. Chadderton. 2004. Ecological and genetic evidence of the vulnerability of shallow-water populations of the stylasterid hydrocoral *Errina novaezelandiae* in New Zealand's fiords. *Aquatic Conservation: Marine and Freshwater Ecosystems* 14: 75-94.
Miller, K. M., K. Laberee, K. H. Kaukinen, S. Li, and R. E. Withler. 2001. Development of microsatellite loci in pinto abalone (*Haliotis kamtschatkana*). *Molecular Ecology Notes* 1: 315-317.
Miller, M. P. 2005. Alleles In Space (AIS): Computer software for the joint analysis of interindividual spatial and genetic information. *Journal of Heredity* 96: 722-724.
Mladenov, P. V., R. M. Allibone, and G. P. Wallis. 1997. Genetic differentiation in the New Zealand sea urchin *Evechinus chloroticus* (Echinodermata: Echinoidea). *New Zealand Journal of Marine and Freshwater Research* 31: 261-269.
Monsen, K. J., and M. S. Blouin. 2003. Genetic structure in a montane ranid frog: restricted gene flow and nuclear-mitochondrial discordance. *Molecular Ecology* 12: 3275-3286.
Moore, L. B. 1961. Distribution patterns of New Zealand seaweeds. *Tuatara (J. Biol. Soc. Vic. Univ. Wellington, NZ)* 9: 18-23.
Moreland, J. 1959. The composition, distribution and origin of New Zealand fish fauna. *Proceedings of the New Zealand Ecological Society* 6: 28-30.

Moritz, C., T. E. Dowling, and W. M. Brown. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annual Review of Ecology and Systematics* 18: 269-292.
Morse, D. E. 1990. Recent progress in larval settlement and metamorphosis: Closing the gaps between molecular biology and ecology. *Bulletin of Marine Science* 46: 465-483.
Morse, D. E. 1992. Molecular mechanisms controlling metamorphosis and recruitment in abalone larvae. Pp. 107-119 *in* S. A. Shepherd, M. J. Tegner, and S. A. Guzman Del Proo, eds. *Abalone of the World: Biology, Fisheries and Culture*. Fishing News Books, Oxford.
Moy, G. W., S. A. Springer, S. L. Adams, W. J. Swanson, and V. D. Vacquier. 2008. Extraordinary intraspecific diversity in oyster sperm bindin. *Proceedings of the National Academy of Sciences of the United States of America* 105: 1993-1998.
Murdoch, R. C., R. Q. Guo, and A. McCrone. 1990. Distribution of Hoki (*Macruronus novaezelandiae*) eggs and larvae in relation to hydrography in eastern Cook Strait, September 1987. *New Zealand Journal of Marine and Freshwater Research* 24: 529-539.
Nagylaki, T. 1998. Fixation indices in subdivided populations *Genetics* 148: 1325-1332.
Naylor, J. R., N. L. Andrew, and S. W. Kim. 2006. Demographic variation in the New Zealand abalone *Haliotis iris*. *Marine and Freshwater Research* 57: 215-224.
Naylor, J. R., and P. E. McShane. 1997. Predation by polychaete worms on larval and post-settlement abalone *Haliotis iris* (Mollusca:Gastropoda). *Journal of Experimental Marine Biology and Ecology* 214: 283-290.
Naylor, J. R., and P. E. McShane. 2001. Mortality of post-settlement abalone *Haliotis iris* caused by conspecific adults and wave exposure. *New Zealand Journal of Marine and Freshwater Research* 35: 363-369.
Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.
Nei, M., and S. Kumar. 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press, Inc., New York.
Nei, M., and W. H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America* 76: 5269-5273.
Neigel, J. E. 2002. Is F-ST obsolete? *Conservation Genetics* 3: 167-173.
Nielsen, R. 2005. Molecular signatures of natural selection. *Annual Review of Genetics* 39: 197-218.
Nikitina, T. V., and S. A. Nazarenko. 2004. Human microsatellites: Mutation and evolution. *Russian Journal of Genetics* 40: 1065-1079.
Niu, T. 2004. Algorithms for inferring haplotypes. *Genetic Epidemiology* 27: 334-347.
Niu, T., Z. S. Qin, X. Xu, and J. S. Liu. 2002. Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms. *American Journal of Human Genetics* 70: 157-169.
Nordborg, N. 2003. Coalescent Theory. Pp. 202-635 *in* D. J. Balding, M. Bishop, and C. Cannings, eds. *Handbook of Statistical Genetics*. Wiley, Chichester, West Sussex, England.
Obbard, D. J., S. A. Harris, and J. R. Pannell. 2006. Simple allelic-phenotype diversity and differentiation statistics for allopolyploids. *Heredity* 97: 296-303.
Officer, R. A., C. Dixon, and H. K. Gorfine. 2001. Movement and re-aggregation of the blacklip abalone, *Haliotis rubra* Leach, after fishing. *Journal of Shellfish Research* 20: 771-779.
Okumura, S. I., S. Furukawa, T. Kawai, S. Takahashi, and K. Yamamori. 2001. Comparison of nucleoli number in diploid and triploid larva of Pacific abalone *Haliotis discus hannai*. *Fisheries Science* 67: 176-178.
Oliveira, E. J., J. G. Padua, M. I. Zucchi, R. Vencovsky, and M. L. C. Vieira. 2006. Origin, evolution and genome distribution of microsatellites. *Genetics and Molecular Biology* 29: 294-307.

Onitsuka, T., T. Kawamura, T. Horii, N. Takiguchi, H. Takami, and Y. Watanabe. 2007. Synchronized spawning of abalone *Haliotis diversicolor* triggered by typhoon events in Sagami Bay, Japan. *Marine Ecology Progress Series* 351: 129-138.

Ostrow, D. G., S. R. Wing, P. V. Mladenov, and M. S. Roy. 2001. Genetic differentiation of *Terebratella sanguinea* in the New Zealand fjords: a dispersal barrier in the marine environment? Pp. 150-159 *in* C. Howard, C. Brunton, L. Robin, M. Cocks, and S. L. Long, eds. *Brachiopods Past and Present*. The Natural History Museum, London.

Ovenden, J. R., D. J. Brasher, and R. W. G. White. 1992. Mitochondrial DNA analyses of the red rock lobster *Jasus edwardsii* supports an apparent absence of population subdivision throughout Australasia. *Marine Biology* 112: 319-326.

Palumbi, S. R. 1992. Marine speciation on a small planet. *Trends in Ecology & Evolution* 7: 114-118.

Palumbi, S. R. 1994. Genetic-divergence, reproductive isolation, and marine speciation. *Annual Review of Ecology and Systematics* 25: 547-572.

Palumbi, S. R. 1999. All males are not created equal: Fertility differences depend on gamete recognition polymorphisms in sea urchins. *Proceedings of the National Academy of Sciences of the United States of America* 96: 12632-12637.

Palumbi, S. R., and A. C. Wilson. 1990. Mitochondrial-DNA diversity in the sea urchins *Strongylocentrotus purpuratus* and *Strongylocentrotus droebachie*. *Evolution* 44: 403-415.

Panchal, M. 2007. The automation of nested clade phylogeographic analysis. *Bioinformatics* 23: 509-510.

Panchal, M., and M. A. Beaumont. 2007. The automation and evaluation of nested clade phylogeographic analysis. *Evolution* 61: 1466-1480.

Panhuis, T. M., N. L. Clark, and W. J. Swanson. 2006. Rapid evolution of reproductive proteins in abalone and Drosophila. *Philosophical Transactions of the Royal Society B-Biological Sciences* 361: 261-268.

Patterson, C. M. 1967. Chromosome number and systematics in stptoneuran snails. *Malacologia* 5: 111-125.

Pawson, D. L. 1961. Distribution patterns of New Zealand echinoderms. *Tuatara (J. Biol. Soc. Vic. Univ. Wellington, NZ)* 9: 9-18.

Peakall, R., and P. E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288-295.

Pearse, D. E., and K. A. Crandall. 2004. Beyond F-ST: Analysis of population genetic data for conservation. *Conservation Genetics* 5: 585-602.

Peijnenburg, K., C. Fauvelot, A. J. Breeuwer, and S. B. J. Menken. 2006. Spatial and temporal genetic structure of the planktonic *Sagitta setosa* (Chaetognatha) in European seas as revealed by mitochondrial and nuclear DNA markers. *Molecular Ecology* 15: 3319-3338.

Perrin, C., S. R. Wing, and M. S. Roy. 2004. Effects of hydrographic barriers on population genetic structure of the sea star *Coscinasterias muricata* (Echinodermata, Asteroidea) in the New Zealand fiords. *Molecular Ecology* 13: 2183-2195.

Perrin, C. M. 2002. The effects of fiord hydrography and environment on the population genetic structures of the sea urchin *Evechinus chloroticus* and the sea star *Coscinasterias muricata* in New Zealand. Ph. D., University of Otago, Dunedin, New Zealand.

Petit, R. J. 2008a. On the falsifiability of the nested clade phylogeographic analysis method. *Molecular Ecology* 17: 1404.

Petit, R. J. 2008b. The coup de grace for the nested clade phylogeographic analysis? *Molecular Ecology* 17: 516-518.

Phillips, N. E., and J. S. Shima. 2006. Differential effects of suspended sediments on larval survival and settlement of New Zealand urchins *Evechinus chloroticus* and abalone *Haliotis iris*. *Marine Ecology Progress Series* 314: 149-158.

Pompanon, F., A. Bonin, E. Bellemain, and P. Taberlet. 2005. Genotyping errors: Causes, consequences and solutions. *Nature Reviews Genetics* 6: 847-859.
Poore, G. C. B. 1972a. Ecology of New Zealand abalones, *Haliotis* species (Mollusca: Gastropoda) 1. Feeding. *New Zealand Journal of Marine and Freshwater Research* 6: 11-22.
Poore, G. C. B. 1972b. Ecology of New Zealand abalones, *Haliotis* species (Mollusca: Gastropoda) 2. Seasonal and diurnal movements. *New Zealand Journal of Marine and Freshwater Research* 6: 246-258.
Poore, G. C. B. 1973. Ecology of New Zealand abalones, *Haliotis* species (Mollusca: Gastropoda) 4. Reproduction. *New Zealand Journal of Marine and Freshwater Research* 7: 67-84.
Pons, O., and K. Chaouche. 1995. Estimation, variance and optimal sampling of gene diversity. 2. Diploid locus. *Theoretical and Applied Genetics* 91: 122-130.
Pons, O., and R. J. Petit. 1995. Estimation, variance and optimal sampling of gene diversity. 1. Haploid locus. *Theoretical and Applied Genetics* 90: 462-470.
Posada, D., and T. R. Buckley. 2004. Model selection and model averaging in phylogenetics: Advantages of akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology* 53: 793-808.
Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.
Posada, D., and K. A. Crandall. 2001. Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology & Evolution* 16: 37-45.
Posada, D., K. A. Crandall, and A. R. Templeton. 2006. Nested clade analysis statistics. *Molecular Ecology Notes* 6: 590-593.
Poulin, E., A. T. Palma, G. Leiva, D. Narvaez, R. Pacheco, S. A. Navarrete, and J. C. Castilla. 2002. Avoiding offshore transport of competent larvae during upwelling events: The case of the gastropod *Concholepas concholepas* in Central Chile. *Limnology and Oceanography* 47: 1248-1255.
Powell, A. W. B. 1955. Mollusca from the Southern Islands of New Zealand. Pp. 1-52. Cape. Exped. Ser. Bull. DSIR, Wellington.
Powell, A. W. B. 1979. *New Zealand Mollusca: Marine, Land and Freshwater Shells*. William Collins Publishers, Ltd., Auckland.
Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.
Rambaut, A. 2002. Se-Al v2.0a11 Carbon. Universty of Oxford.
Ray, N., M. Currat, and L. Excoffier. 2003. Intra-deme molecular diversity in spatially expanding populations. *Molecular Biology and Evolution* 20: 76-86.
Raymond, M., and F. Rousset. 1995. An exact test for population differentiation. *Evolution* 49: 1280-1283.
Reid, D. G., K. Lal, J. Mackenzie-Dodds, F. Kaligis, D. T. J. Littlewood, and S. T. Williams. 2006. Comparative phylogeography and species boundaries in *Echinolittorina* snails in the central Indo-West Pacific. *Journal of Biogeography* 33: 990-1006.
Reynolds, J., B. S. Weir, and C. C. Cockerham. 1983. Estimation for the coancestry coefficient: basis for a short-term genetic distance. *Genetics* 105: 767-779.
Reynolds-Fleming, J. V., and J. G. Fleming. 2005. Coastal circulation within the Banks Peninsula region, New Zealand. *New Zealand Journal of Marine and Freshwater Research* 39: 217-225.
Richards, V. P., J. D. Thomas, M. J. Stanhope, and M. S. Shivji. 2007. Genetic connectivity in the Florida reef system: comparative phylogeography of commensal invertebrates with contrasting reproductive strategies. *Molecular Ecology* 16: 139-157.
Riginos, C., and J. H. McDonald. 2003. Positive selection on an acrosomal sperm protein, M7 lysin, in three species of the mussel genus *Mytilus*. *Molecular Biology and Evolution* 20: 200-207.

Riginos, C., D. Wang, and A. J. Abrams. 2006. Geographic variation and positive selection on M7 lysin, an acrosomal sperm protein in mussels (*Mytilus* spp.). *Molecular Biology and Evolution* 23: 1952-1965.

Roberts, P. E., and L. J. Paul. 1978. Seasonal hydrological changes in continental shelf waters of the west coast, North Island, New Zealand, and comments on fish distributions. *New Zealand Journal of Marine and Freshwater Research* 12: 323-339.

Roberts, R. 2001. A review of settlement cues for larval abalone (*Haliotis* spp.). *Journal of Shellfish Research* 20: 571-586.

Roberts, R. D., H. F. Kaspar, and R. J. Barker. 2004. Settlement of abalone (*Haliotis* iris) larvae in response to five species of coralline algae. *Journal of Shellfish Research* 23: 975-987.

Roberts, R. D., E. F. Keys, G. Prendeville, and C. A. Pilditch. 2007. Viability of abalone (*Haliotis iris*) stock enhancement by release of hatchery-reared seed in Marlborough, New Zealand. *Journal of Shellfish Research* 26: 697-703.

Roberts, R. D., and C. Lapworth. 2001. Effect of delayed metamorphosis on larval competence, and post-larval survival and growth, in the abalone *Haliotis iris* Gmelin. *Journal of Experimental Marine Biology and Ecology* 258: 1-13.

Rodrígues, F., O. A. Marín, and J. R. Medina. 1990. The general stochastic model of nucleotide substitution. *Journals of Theoretical Biology* 142: 485-501.

Roff, D. A., and P. Bentzen. 1989. The statistical analysis of mitochondrial DNA polymorphisms: χ^2 and the problem of small samples. *Molecular Biology and Evolution* 6: 539-545.

Rogers, A. R., and H. Harpending. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* 9: 552-569.

Rogers-Bennett, L., R. F. Dondanville, and J. Kashiwada. 2004. Size specific fecundity of red abalone (*Haliotis rufescens*): Evidence for reproductive senescence? *Journal of Shellfish Research* 23: 553-560.

Rogers-Bennett, L., D. W. Rogers, and S. A. Schultz. 2007. Modeling growth and mortality of red abalone (*Haliotis rufescens*) in Northern California. *Journal of Shellfish Research* 26: 719-727.

Rokas, A., and S. B. Carroll. 2005. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. *Molecular Biology and Evolution* 22: 1337-1344.

Roughgarden, J., S. Gaines, and H. Possingham. 1988. Recruitment Dynamics in Complex Life-Cycles. *Science* 241: 1460-1466.

Rousset, F. 2008. GENEPOP ' 007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* 8: 103-106.

Rozen, S., and H. J. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. Pp. 365-386 *in* S. Krawetz, and S. Misener, eds. *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ.

Sainsbury, K. J. 1982a. Population dynamics and fishery management of the Paua, *Haliotis iris* .1. Population structure, growth, reproduction, and mortality. *New Zealand Journal of Marine and Freshwater Research* 16: 147-161.

Sainsbury, K. J. 1982b. Population dynamics and fishery management of the paua, *Haliotis iris* .2. Dynamics and management as examined using a size class population model. *New Zealand Journal of Marine and Freshwater Research* 16: 163-173.

Saitou, N., and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4: 406-425.

Sasaki, R., and S. A. Shepherd. 2001. Ecology and post-settlement survival of the ezo abalone, *Haliotis discus hannai*, on Miyagi coasts, Japan. *Journal of Shellfish Research* 20: 619-626.

Scheet, P., and M. Stephens. 2006. A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. *American Journal of Human Genetics* 78: 629-644.

Schiel, D. R. 1993. Experimental evaluation of commercial-scale enhancement of abalone *Haliotis iris* populations in New Zealand. *Marine Ecology Progress Series* 97: 167-181.
Schiel, D. R., and P. A. Breen. 1991. Population structure, ageing, and fishing mortality of the New Zealand abalone *Haliotis iris*. *Fishery Bulletin* 89: 681-691.
Schlotterer, C. 2000. Evolutionary dynamics of microsatellite DNA. *Chromosoma* 109: 365-371.
Schnabel, K. E., I. D. Hogg, and M. A. Chapman. 2000. Population genetic structures of two New Zealand corophiid amphipods and the presence of morphologically cryptic species: implications for the conservation of diversity. *New Zealand Journal of Marine and Freshwater Research* 34: 637-644.
Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* 18: 233-234.
Scribner, K. T., M. R. Petersen, R. L. Fields, S. L. Talbot, J. M. Pearce, and R. K. Chesser. 2001. Sex-biased gene flow in spectacled eiders (anatidae): Inferences from molecular markers with contrasting modes of inheritance. *Evolution* 55: 2105-2115.
Searle, T., R. D. Roberts, and P. M. Lokman. 2006. Effects of temperature on growth of juvenile blackfoot abalone, *Haliotis iris* Gmelin. *Aquaculture Research* 37: 1441-1449.
Sekino, M., and M. Hara. 2001. Microsatellite DNA loci in Pacific abalone *Haliotis discus discus* (Mollusca, Gastropoda, Haliotidae). *Molecular Ecology Notes* 1: 8-10.
Sekino, M., T. Kobayashi, and M. Hara. 2006. Segregation and linkage analysis of 75 novel microsatellite DNA markers in pair crosses of Japanese abalone (*Haliotis discus hannai*) using the 5 '-tailed primer method. *Marine Biotechnology* 8: 453-466.
Sekino, M., T. Saido, T. Fujita, T. Kobayashi, and H. Takami. 2005. Microsatellite DNA markers of Ezo abalone (*Haliotis discus hannai*): A preliminary assessment of natural populations sampled from heavily stocked areas. *Aquaculture* 243: 33-47.
Selkoe, K. A., and R. J. Toonen. 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters* 9: 615-629.
Selvamani, M. J. P., S. M. Degnan, and B. M. Degnan. 2001. Microsatellite genotyping of individual abalone larvae: Parentage assignment in aquaculture. *Marine Biotechnology* 3: 478-485.
Selvamani, M. J. P., S. M. Degnan, D. Paetkau, and B. M. Degnan. 2000. Highly polymorphic microsatellite loci in the Heron Reef population of the tropical abalone *Haliotis asinina*. *Molecular Ecology* 9: 1184-1186.
Serviere-Zaragoza, E., A. Mazariegos-Villareal, G. Ponce-Diaz, and S. M. Magallon. 2001. Growth of juvenile abalone, *Haliotis fulgens* philippi, fed different diets. *Journal of Shellfish Research* 20: 689-693.
Shanks, A. L., and L. Brink. 2005. Upwelling, downwelling, and cross-shelf transport of bivalve larvae: test of a hypothesis. *Marine Ecology Progress Series* 302: 1-12.
Shaw, A., P. A. G. Fortes, C. D. Stout, and V. D. Vacquier. 1995. Crystal-structure and subunit dynamics of the abalone sperm lysin dimer: Egg envelopes dissociate dimers, the monomer is the active species. *Journal of Cell Biology* 130: 1117-1125.
Shepherd, S. A. 1986. Studies on southern Australian abalone (Genus *Haliotis*) .7. Aggregative behavior of *H. laevigata* in relation to spawning. *Marine Biology* 90: 231-236.
Shepherd, S. A., J. L. Baker, and D. W. Johnson. 1995. Yield-per-recruit and egg-per-recruit analyses of the Omani abalone, *Haliotis mariae*. *Marine and Freshwater Research* 46: 663-668.
Shepherd, S. A., D. Lowe, and D. Partington. 1992. Studies on southern Australian abalone (Genus *Haliotis*) .13. Larval dispersal and recruitment. *Journal of Experimental Marine Biology and Ecology* 164: 247-260.
Shilling, F. M., O. HoeghGuldberg, and D. T. Manahan. 1996. Sources of energy for increased metabolic demand during metamorphosis of the abalone *Haliotis rufescens* (Mollusca). *Biological Bulletin* 191: 402-412.

Shirtcliffe, T. G. L., M. I. Moore, A. G. Cole, A. B. Viner, R. Baldwin, and B. Chapman. 1990. Dynamics of the Cape Farewell upwelling plume, New Zealand. *New Zealand Journal of Marine and Freshwater Research* 24: 555-568.

Simonsen, K. L., G. A. Churchill, and C. F. Aquadro. 1995. Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* 141: 413-429.

Sköld, M., S. R. Wing, and P. V. Mladenov. 2003. Genetic subdivision of a sea star with high dispersal capability in relation to physical barriers in a fjordic seascape. *Marine Ecology Progress Series* 250: 163-174.

Slatkin, M. 1985. Gene flow in natural populations. *Annual Review in Ecology and Systematics* 16: 393-430.

Slatkin, M. 1991. Inbreeding coefficients and coalescence times. *Genetics Research Cambridge* 58.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139: 457-462.

Slatkin, M., and R. R. Hudson. 1991. Pairwise comparisons of mitochondrial-DNA sequences in stable and exponentially growing populations. *Genetics* 129: 555-562.

Small, K. S., M. Brudno, M. M. Hill, and A. Sidow. 2007. Extreme genomic variation in a natural population. *Proceedings of the National Academy of Sciences of the United States of America* 104: 5698-5703.

Smith, F. 2001. Historical regulation of local species richness across a geographic region. *Ecology* 82: 792-801.

Smith, J. R., J. D. Carpten, M. J. Brownstein, S. Ghosh, V. L. Magnuson, D. A. Gilbert, J. M. Trent, and F. S. Collins. 1995. Approach to genotyping errors caused by nontemplated nucleotide addition by Taq DNA polymerase. *Genome Research* 5: 312-317.

Smith, P. J. 1988. Biochemical-genetic variation in the green-lipped mussel *Perna canaliculus* around New Zealand and possible implications for mussel farming. *New Zealand Journal of Marine and Freshwater Research* 22: 85-90.

Smith, P. J., and P. G. Benson. 1997. Genetic diversity in orange roughy from the east of New Zealand. *Fisheries Research* 31: 197-213.

Smith, P. J., P. G. Benson, and S. M. McVeagh. 1997. A comparison of three genetic methods used for stock discrimination of orange roughy, *Hoplostethus atlanticus*: allozymes, mitochondrial DNA, and random amplified polymorphic DNA. *Fishery Bulletin* 95: 800-811.

Smith, P. J., and A. M. Conroy. 1992. Loss of genetic-variation in hatchery-produced abalone, *Haliotis iris*. *New Zealand Journal of Marine and Freshwater Research* 26: 81-85.

Smith, P. J., G. J. Macarthur, and K. P. Michael. 1989. Regional variation in electromorph frequencies in the tuatua, *Paphies subtriangulata*, around New Zealand. *New Zealand Journal of Marine and Freshwater Research* 23: 27-33.

Smith, P. J., J. L. McKoy, and P. J. Machin. 1980. Genetic variation in the rock lobsters *Jasus edwardsii* and *Jasus novaehollandiae*. *New Zealand Journal of Marine and Freshwater Research* 14: 55 - 63.

Smith, P. J., and M. S. McVeagh. 2006. SAP2005-01: Genetic population structure of blackfoot paua. N. Z. M. o. Fisheries.

Smith, P. J., H. Ozaki, and Y. Fujio. 1986. No evidence for reduced genetic variation in accidentally introduced oyster *Crassostrea gigas* in New Zealand *New Zealand Journal of Marine and Freshwater Research* 20: 569-574.

Sokal, R. R., and D. E. Wartenberg. 1983. A test of spatial autocorrelation analysis using an isolation by distance model. *Genetics* 105: 219-237.

Sponaugle, S., R. K. Cowen, A. Shanks, S. G. Morgan, J. M. Leis, J. S. Pineda, G. W. Boehlert, M. J. Kingsford, K. C. Lindeman, C. Grimes, and J. L. Munro. 2002. Predicting self-recruitment in marine populations: Biophysical correlates and mechanisms. *Bulletin of Marine Science* 70: 341-375.

Sponer, R., and M. S. Roy. 2002. Phylogeographic analysis of the brooding brittle star *Amphipholis squamata* (Echinodermata) along the coast of New Zealand reveals high cryptic genetic variation and cryptic dispersal potential. *Evolution* 56: 1954-1967.
Springer, S. A., and B. J. Crespi. 2007. Adaptive gamete-recognition divergence in a hybridizing *Mytilus* population. *Evolution* 61: 772-783.
Stanton, B. R. 1973. Hydrological investigations around northern New Zealand. *New Zealand Journal of Marine and Freshwater Research* 3: 124-146.
Stanton, B. R., P. J. H. Sutton, and S. M. Chiswell. 1997. The East Auckland Current, 1994-95. *New Zealand Journal of Marine and Freshwater Research* 31: 537-549.
Star, B., S. Apte, and J. P. A. Gardner. 2003. Genetic structuring among populations of the greenshell mussel *Perna canaliculus* revealed by analysis of randomly amplified polymorphic DNA. *Marine Ecology Progress Series* 249: 171-182.
Steinarsson, A., and A. K. Imsland. 2003. Size dependent variation in optimum growth temperature of red abalone (*Haliotis rufescens*). *Aquaculture* 224: 353-362.
Stephens, M., and P. Donnelly. 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics* 73: 1162-1169.
Stephens, M., and P. Scheet. 2005. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *American Journal of Human Genetics* 76: 449-462.
Stephens, M., N. J. Smith, and P. Donnelly. 2001. A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* 68: 978-989.
Stephens, S. A., N. Broekhuizen, A. B. Macdiarmid, C. J. Lundquist, L. McLeod, and R. Haskew. 2006. Modelling transport of larval New Zealand abalone (*Haliotis iris*) along an open coast. *Marine and Freshwater Research* 57: 519-532.
Stevens, G. R., M. McGlone, and B. McCulloch. 1995. *Prehistoric New Zealand*. Reed Publishing (NZ) Ltd., Auckland.
Stevens, M. I., and I. D. Hogg. 2004. Population genetic structure of New Zealand's endemic corophiid amphipods: evidence for allopatric speciation. *Biological Journal of the Linnean Society* 81: 119-133.
Stevens, P. M. 1991. A genetic analysis of the pea crabs (Decapoda, Pinnotheridae) of New Zealand. 2. Patterns and intensity of spaction population structure in *Pinnotheres atrinicola*. *Marine Biology* 108: 403-410.
Streit, K., D. L. Geiger, and B. Lieb. 2006. Molecular phylogeny and the geographic origin of haliotidae traced by haemocyanin sequences. *Journal of Molluscan Studies* 72: 105-110.
Stuart, M. D., and M. T. Brown. 1994. Growth and diet of cultivated black-footed abalone, *Haliotis iris* (Martyn). *Aquaculture* 127: 329-337.
Sun, X. Q., M. G. Zhen, and G. P. Yang. 2007. Development of 15 polymorphic genic microsatellite DNA markers of Pacific abalone *Haliotis discus hannai*. *Molecular Ecology Notes* 7: 604-606.
Sunnucks, P. 2000. Efficient genetic markers for population biology. *Trends in Ecology & Evolution* 15: 199-203.
Swanson, W. J., C. F. Aquadro, and V. D. Vacquier. 2001. Polymorphism in abalone fertilization proteins is consistent with the neutral evolution of the egg's receptor for lysin (VERL) and positive Darwinian selection of sperm lysin. *Molecular Biology and Evolution* 18: 376-383.
Swanson, W. J., and V. D. Vacquier. 1997. The abalone egg vitelline envelope receptor for sperm lysin is a giant multivalent molecule. *Proceedings of the National Academy of Sciences of the United States of America* 94: 6724-6729.
Swanson, W. J., and V. D. Vacquier. 1998. Concerted evolution in an egg receptor for a rapidly evolving abalone sperm protein. *Science* 281: 710-712.
Swanson, W. J., and V. D. Vacquier. 2002. The rapid evolution of reproductive proteins. *Nature Reviews Genetics* 3: 137-144.

Swofford, D. L. 1998. PAUP*: Phylogenetic Analysis Using Parsimony (*and other methods). Sinauer Associates.
Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585-595.
Takagi, T., A. Nakamura, R. Deguchi, and K. Kyozuka. 1994. Isolation, characterization, and primary structure of 3 major proteins obtained from *Mytilus edulis* sperm. *Journal of Biochemistry* 116: 598-605.
Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 24: 1596-1599.
Tamura, K., and M. Nei. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial-DNA in humans and chimpanzees. *Molecular Biology and Evolution* 10: 512-526.
Tang, S., A. Popongviwat, S. Klinbunga, A. Tassanakajon, P. Jarayabhand, and P. Menasveta. 2005. Genetic heterogeneity of the tropical abalone (*Haliotis asinina*) revealed by RAPD and Microsatellite analyses. *Journal of Biochemistry and Molecular Biology* 38: 182-190.
Tang, S., A. Tassanakajon, S. Klinbunga, P. Jarayabhand, and P. Menasveta. 2004. Population structure of tropical abalone (*Haliotis asinina*) in coastal waters of Thailand determined using microsatellite markers. *Marine Biotechnology* 6: 604-611.
Tarr, R. J. Q. 1995. Growth and movement of the South African albalone *Haliotis midae* - a reassessment. *Marine and Freshwater Research* 46: 583-590.
Tatusov, R. L., E. V. Koonin, and D. J. Lipman. 1997. A genomic perspective on protein families. *Science* 278: 631-637.
Tavaré, S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences *in* R. M. Miura, ed. *Some mathematical questions is biology - DNA sequence analysis*. American Mathematical Society, Providence, RI, USA.
Team, R. C. D. 2007. A language and environment for statistical computing. R Foundation for Statistical Computing.
Temby, N., K. Miller, and C. Mundy. 2007. Evidence of genetic subdivision among populations of blacklip abalone (*Haliotis rubra Leach*) in Tasmania. *Marine and Freshwater Research* 58: 733-742.
Templeton, A. R. 1998. Nested clade analyses of phylogeographic data: Testing hypotheses about gene flow and population history. *Molecular Ecology* 7: 381-397.
Templeton, A. R. 2004. Statistical phylogeography: methods of evaluating and minimizing inference errors. *Molecular Ecology* 13: 789-809.
Templeton, A. R., E. Boerwinkle, and C. F. Sing. 1987. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping.1. Basic theory and an analysis of alcohol-dehydrogenase activity in drosophila. *Genetics* 117: 343-351.
Templeton, A. R., K. A. Crandall, and C. F. Sing. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence Data. 3. Cladogram estimation. *Genetics* 132: 619-633.
Templeton, A. R., E. Routman, and C. A. Phillips. 1995. Separating population structure from population history: A cladistic analysis of the geographical distribution of mitochondrial-DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics* 140: 767-782.
Templeton, A. R., and C. F. Sing. 1993. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. 4. Nested analyses with cladogram uncertainty and recombination. *Genetics* 134: 659-669.
Thiel, M., and L. Gutow. 2005. The ecology of rafting in the marine environment. II. The rafting organisms and community. Pp. 279-418 *in* R. N. Gibson, R. J. A. Atkinson, and J. D. M. Gordon, eds. *Oceanography and Marine Biology - an Annual Review, Vol. 43*. Taylor & Francis Group, New York.

Tissot, B. N. 1992. Water movement and the ecology and evolution of the Haliotidae. Pp. 34-45 *in* S. A. Shepherd, M. J. Tegner, and S. A. G. del Proo, eds. *Abalone of the World: Biology, Fisheries and Culture*. Fishing News Books, Oxford.

Todd, C. D. 1998. Larval supply and recruitment of benthic invertebrates: Do larvae always disperse as much as we believe? *Hydrobiologia* 375-376: 1-21.

Todd, C. D., W. J. Lambert, and J. P. Thorpe. 1998. The genetic structure of intertidal populations of two species of nudibranch molluscs with planktotrophic and pelagic lecithotrophic larval stages: Are pelagic larvae 'for' dispersal? *Journal of Experimental Marine Biology and Ecology* 228: 1-28.

Tong, L. J., G. A. Moss, P. Redfearn, and J. Illingworth. 1992. A manual of techniques for culturing paua, *Haliotis iris*, through to the early juvenile stage.

Tóth, G., Z. Gáspári, and J. Jurka. 2000. Microsatellites in different eukaryotic genomes: Surveys and analysis. *Genome Research* 10: 967-981.

Uddstrom, M. J., and N. A. Oien. 1999. On the use of high-resolution satellite data to describe the spatial and temporal variability of sea surface temperatures in the New Zealand region. *Journal of Geophysical Research-Oceans* 104: 20729-20751.

Underwood, A. J., and P. G. Fairweather. 1989. Supply-side ecology and benthic marine assemblages. *Trends in Ecology & Evolution* 4: 16-20.

Uthicke, S., and J. A. H. Benzie. 2003. Gene flow and population history in high dispersal marine invertebrates: mitochondrial DNA analysis of *Holothuria nobilis* (Echinodermata : Holothuroidea) populations from the Indo-Pacific. *Molecular Ecology* 12: 2635-2648.

Vacquier, V. D., K. R. Carner, and C. D. Stout. 1990. Species-specific sequences of abalone lysin, the sperm protein that creates a hole in the egg envelope. *Proceedings of the National Academy of Sciences of the United States of America* 87: 5792-5796.

Vacquier, V. D., W. J. Swanson, and M. E. Hellberg. 1995. What have we learned about sea urchin sperm bindin. *Development Growth & Differentiation* 37: 1-10.

Van Oosterhout, C., W. F. Hutchinson, D. P. M. Wills, and P. Shipley. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4: 535-538.

Veale, A. J. 2007. Phylogeography of two intertidal benthic marine invertebrates around New Zealand, the waratah anemone (*Actinia tenebrosa*) and the snakeskin chiton (*Sypharochiton pelliserpentis*). M. Sc., Univerisity of Auckland, Auckland.

Venables, W. N., and B. D. Ripley. 1999. *Modern Applied Statistics with S-PLUS*. Springer-Verlag New York, Inc., New York, USA.

Vincent, W. F., C. Howard-Williams, P. Tildesley, and E. Butler. 1991. Distribution and biological properties of oceanic water masses around the South Island, New Zealand. *New Zealand Journal of Marine and Freshwater Research* 25: 21-42.

Wan, Q. H., H. Wu, T. Fujihara, and S. G. Fang. 2004. Which genetic marker for which conservation genetics issue? *Electrophoresis* 25: 2165-2176.

Wang, A., P. Roffey, and C. Blanchard. 2006. Differentiation of Australian and New Zealand abalone species based on partial mitochondrial gene sequences of 12S rRNA, 16S rRNA and ND2 (NADH dehydrogenase subunit 2). *Molluscan Research* 26: 98-102.

Wang, L., and Y. Xu. 2003. Haplotype inference by maximum parsimony. *Bioinformatics* 19: 1773-1780.

Wang, L. L., H. Zhang, L. S. Song, and X. M. Guo. 2007. Loss of allele diversity in introduced populations of the hermaphroditic bay scallop *Argopecten irradians*. *Aquaculture* 271: 252-259.

Wares, J. P., S. D. Gaines, and C. W. Cunningham. 2001. A comparative study of asymmetric migration events across a marine biogeographic boundary. *Evolution* 55: 295-306.

Waters, J. M., T. M. King, P. M. O'Loughlin, and H. G. Spencer. 2005. Phylogeographical disjunction in abundant high-dispersal littoral gastropods. *Molecular Ecology* 14: 2789-2802.

Waters, J. M., and M. S. Roy. 2004. Phylogeography of a high-dispersal New Zealand sea star: Does upwelling block gene-flow? *Molecular Ecology* 13: 2797-2806.

Watts, P. C., J. P. Thorpe, and P. D. Taylor. 1998. Natural and anthropogenic dispersal mechanisms in the marine environment: a study using cheilostome Bryozoa. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 353: 453-464.

Weir, B. S., and W. G. Hill. 2002. Estimating F-statistics. *Annual Review of Genetics* 36: 721-750.

Wells, F. E., and P. Mulvay. 1995. Good and bad fishing areas for *Haliotis laevigata*: A comparison of population parameters. *Marine and Freshwater Research* 46: 591-598.

Werner, I., S. Flothmann, and G. Burnell. 1995. Behavior studies on the mobility of 2 Species of abalone (*Haliotis tuberculata* and *Haliotis discus hannai*) on sand: Implications for reseeding programs. *Marine and Freshwater Research* 46: 681-688.

Whitlock, M. C., and D. E. McCauley. 1999. Indirect measures of gene flow and migration: F-ST not equal 1/(4Nm+1). *Heredity* 82: 117-125.

Wilhelm, R., and T. J. Hilbish. 1998. Assessment of natural selection in a hybrid population of mussels: Evaluation of exogenous vs endogenous selection models. *Marine Biology* 131: 505-514.

Williams, S. T., and T. Ozawa. 2006. Molecular phylogeny suggests polyphyly of both the turban shells (family Turbinidae) and the superfamily Trochoidea (Mollusca : Vetigastropoda). *Molecular Phylogenetics and Evolution* 39: 33-51.

Willis, G. L., and D. O. F. Skibinski. 1992. Variation in strength of attachment to the substrate explains differential mortality in hybrid mussel (*Mytilus galloprovincialis* and *M. edulis*) populations *Marine Biology* 112: 403-408.

Winnepenninckx, B., G. Steiner, T. Backeljau, and R. De Wachter. 1998. Details of gastropod phylogeny inferred from 18S rRNA sequences. *Molecular Phylogenetics and Evolution* 9: 55-63.

Withler, R. E. 2000. Genetic tools for identification and conservation of exploited abalone (*Haliotis* spp.) species. Pp. 101-110 *in* A. Campbell, ed. *Workshop on Rebuilding Abalone Stocks in British Columbia*. NRC Research Press, Ottawa.

Withler, R. E., A. Campbell, S. Li, K. M. Miller, D. Brouwer, and B. G. Lucas. 2001. High Levels of Genetic Variation in Northern Abalone *Haliotis kamtschatkana* of British Colombia.

Withler, R. E., A. Campbell, S. R. Li, D. Brouwer, K. J. Supernault, and K. M. Miller. 2003. Implications of high levels of genetic diversity and weak population structure for the rebuilding of northern abalone in British Columbia, Canada. *Journal of Shellfish Research* 22: 839-847.

Wodicka, L. M., and D. E. Morse. 1991. cDNA sequences reveal messenger RNAs for 2 G-alpha signal transducing proteins from larval cilia. *Biological Bulletin* 180: 318-327.

Wolfe, K. H., W.-H. Li, and P. M. Sharp. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences of the United States of America* 84: 9054-9058.

Won, Y., C. R. Young, R. A. Lutz, and R. C. Vrijenhoek. 2003. Dispersal barriers and isolation among deep-sea mussel populations (Mytilidae: *Bathymodiolus*) from eastern Pacific hydrothermal vents. *Molecular Ecology* 12: 169-184.

Wong, W. S. W., and R. Nielsen. 2004. Detecting selection in noncoding regions of nucleotide sequences. *Genetics* 167: 949-958.

Wood, A. D., and C. D. Buxton. 1996. Aspects of the biology of the abalone *Haliotis midae* (Linne, 1758) on the east coast of South Africa. 2. Reproduction. *South African Journal of Marine Science-Suid-Afrikaanse Tydskrif Vir Seewetenskap* 17: 69-78.

Worcester, S. E. 1994. Adult rafting versus larval swimming - dispersal and recruitment of a

botryllid ascidian on eel grass. *Marine Biology* 121: 309-317.
Worthington, D. G., and N. L. Andrew. 1997. Does covariation between growth and reproduction compromise the use of an alternative size limit for the blacklip abalone, *Haliotis rubra*, in NSW, Australia? *Fisheries Research* 32: 223-231.
Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16.
Xu, J. P. 2005. The inheritance of organelle genes and genomes: patterns and mechanisms. *Genome* 48: 951-958.
Yang, H.-S., Y.-Y. Ting, and H.-C. Chen. 1998a. Blocking polar body with cytochalasin B in the fertilized eggs of the small abalone, *Haliotis diversicolor supertexta* (Lischke), and the development and ploidy of the resultant embryos. *Aquaculture Research* 29: 775-783.
Yang, H. S., H. C. Chen, and Y. Y. Ting. 1998b. Induction of polyploidy and embryonic development of the abalone, *Haliotis diversicolor*, with temperature treatment. *American Malacological Bulletin* 14: 139-147.
Yang, Z. H., and W. J. Swanson. 2002. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. *Molecular Biology and Evolution* 19: 49-57.
Yang, Z. H., W. J. Swanson, and V. D. Vacquier. 2000. Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. *Molecular Biology and Evolution* 17: 1446-1455.
Zane, L., L. Bargelloni, and T. Patarnello. 2002. Strategies for microsatellite isolation: a review. *Molecular Ecology* 11: 1-16.
Zhang, D. X., and G. M. Hewitt. 2003. Nuclear DNA analyses in genetic studies of populations: Practice, problems and prospects. *Molecular Ecology* 12: 563-584.
Zhang, S., A. J. Pakstis, K. K. Kidd, H. Zhao, M. Stephens, N. J. Smith, and P. Donnelly. 2001. Comparisons of two methods for haplotype reconstruction and haplotype frequency estimation from population data [2]. *American Journal of Human Genetics* 69: 906-914.
Zhang, Y., T. Niu, and J. S. Liu. 2006. A coalescence-guided hierarchical Bayesian method for haplotype inference. *American Journal of Human Genetics* 79: 313-322.
Zhang, Z., A. Campbell, and J. Lessard. 2007. Modeling northern abalone, *Haliotis kamtschatkana*, population stock and recruitment in British Columbia. *Journal of Shellfish Research* 26: 1099-1107.
Zigler, K. S., and H. A. Lessios. 2004. Speciation on the coasts of the new world: Phylogeography and the evolution of bindin in the sea urchin genus *Lytechinus*. *Evolution* 58: 1225-1241.
Zigler, K. S., M. A. McCartney, D. R. Levitan, and H. A. Lessios. 2005. Sea urchin bindin divergence predicts gamete compatibility. *Evolution* 59: 2399-2404.
Zink, R. M., and G. F. Barrowclough. 2008. Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology* 17: 2107-2121.
Zouros, E. 2000. The exceptional mitochondrial DNA system of the mussel family Mytilidae. *Genes & Genetic Systems* 75: 313-318.
Zúñiga, G., S. A. Guzmán del Próo, R. Cisneros, and G. Rodríguez. 2000. Population genetic analysis of the abalone *Haliotis fulgens* (Mollusca : Gastropoda) in Baja California, Mexico. *Journal of Shellfish Research* 19: 853-859.

Appendices

APPENDIX 1: SAMPLES

Table 1: Details for *Haliotis iris* samples currently held in the Molecular Ecology Laboratory, University of Otago, Dunedin, New Zealand.

Location	ID	N	Tissue type	Collector	Coordinates*	Collection Date
Ahu Ahu Road	AHU	22	whole	Ministry of Fisheries	-39.117159° 173.929819°	16 Oct 2005
Cape Campbell	CBL	14	whole	UC	-41.741920° 174.275685° -41.734349° 174.276258° -41.725371° 174.273675°	15 Oct 2005
Colac Bay	CCB	20	cast-off	Riverton Fisheries, LTD.	-46.422131° 167.836311°	13 Jan 2005
Lottin Point, Cape Runaway	CRW	20	foot and epipodia	Ministry of Fisheries	-37.547890° 178.166130°	2 Nov 2005
Doubtless Bay	DBL	15	whole	Department of Conservation	-34.849000° 173.470007°	2 May 2005
Doubtful Sound	DSD	14	cast-off	Riverton Fisheries, LTD.	-45.269575° 166.889359°	12 Jan 2005
East Island	EAI	24	cast-off	Ministry of Fisheries	-37.690435° 178.577756°	19 Dec 05
Raglan	GLN	20	whole	Professional fisherman	-37.820013° 173.801865°	26 Jun 2006
Goose Bay	GOB	19	cast-off	Recreational fisherman	-42.482263° 173.529256°	8 Jan 05
Pihama	IHM	20	whole	Ministry of Fisheries	-39.521503° 173.913828°	16 Oct 05
Cascade Point	JCH	20	whole	Professional fisherman	-44.008338° 168.365705°	6 Jul 07
Castle Point	MAT	21	cast-off	Top Cat Abalone & Venison Products	-40.882115° 176.224620°	18 Jan 2005
Magnet Bay	MTB	22	whole	UC	-43.841655° 172.738753°	6 Feb 05
Nugget Point	NPT	20	cast-off	Riverton, LTD.	-46.481505° 169.755918°	13 Jan 05
Owenga, Chatham Island	OCH	13	whole	Ministry of Fisheries	-45.005556° 176.455556	6 Dec 05
Tolaga Bay	OLB	20	whole	Ministry of Fisheries	-38.378430° 178.342005°	19 Jan 06
Opito	OPT	20	whole	UC	-36.716311° 175.817936°	10 Apr 06
Port Hardy	PHD	20	cast-off	Burkheart Fisheries	-40.750326° 173.887572°	29 Nov 05
Spirits Bay	SPB	21	whole	UC	-34.417460° 172.85574°	16 Apr 06
Ruggedy, Stewart Island	STR	15	cast-off	Riverton, LTD.	-46.705408° 167.716418°	14 Nov 06
Tory Channel	TCL	20	whole	Professional fisherman	-41.205445° 174.305533°	27 Sep 06
Timaru	TIM	23	whole	UC	-44.375875° 171.252545°	7 Jan 06
Taylor's Mistake	TSK	20	whole	UC	-43.585165° 172.789056°	11 Dec 05
Wellington	WLG	20	whole	Professional fisherman	-41.337139° 174.792826°	19 Jul 06
West Haven	WST	15	cast-off	Burkheart Fisheries	-40.565063° 172.553310°	29 Nov 05
TOTAL		**478**				

*Coordinates for samples collected are approximations.

APPENDIX 2: MITOCHONDRIAL HAPLOTYPES

Listed below is the alignment in FASTA format for the 132 haplotypes for concatenated mtCOI and ATP8 – ATP6. Labels refer to the numbered haplotypes in Figures 2.5–2.7.

```
>h0
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATACATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h1
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCACAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h2
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGGCCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h3
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
```

AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCCCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h4
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATCCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCGCCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h5
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h6
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAAAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCCCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h7
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC

GTAAAAATTACCGCCGTCCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h8
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h9
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTGGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h10
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h11
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCTGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG

CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCCCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGATCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTTATAGCAGCACTATT
TACCTCCCTATTAGT

>h12
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCTGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCCCTCTTTTAAGGATTAAGATTTAAGTCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGATCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTTATAGCAGCACTATT
TACCTCCCTATTAGT

>h13
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGATGAACAGTTTATCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGAGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h14
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGACCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h15
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG

TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCCCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGATCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h16
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGATGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h17
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGGCCTTGGTAACTAAACACACCATTACTTAAATTAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAGACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTGTT
TACCTCCCTATTAGT

>h18
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGATAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h19
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA

TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCCTGGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGGCCTTGGTAACTAAACACACCATTACTTAAATTAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAGACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTGTT
TACCTCCCTATTAGT

>h20
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCGTCCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h21
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGATGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATCTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h22
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCGCTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h23
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT

GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCTTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CGCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h24
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCCCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAATTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h25
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTGGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGATCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTTATAGCAGCACTATT
TACCTCCCTATTAGT

>h26
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCATAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h27

CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGATGAACAGTTTACCCTCCTTTATCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCACTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCACACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h28
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAGCACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTCCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGATCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTTATAGCAGCACTATT
TACCTCCCTATTAGT

>h29
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGATGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCACACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h30
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTGACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGATGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCACACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h31
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCGTCCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCCCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h32
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCACTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGGCCTTGGTAACTAAACACACCATTACTTAAATTAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAGACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTGTT
TACCTCCCTATTAGT

>h33
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h34
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCCCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCTAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAATCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h35
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTATCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTACTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h36
CAATGTAATTGTAACAGCCCATGCTTTTGTCATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h37
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGATGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGAGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h38
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATACTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT

TACCTCCCTATTAGT

>h39
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAGCCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h40
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACCTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h41
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTGTCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h42
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT

AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h43
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATATAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h44
CAATGTAATTGTAACAGCCCATGCTTTTGTAATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h45
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-GCCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAACAGCACTATT
TACCTCCCTATTAGT

>h46
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTGGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT

GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h47
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAACGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGGCCTTGGTAACTAAACACACCATTACTTAAATTAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAGACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTGTT
TACCTCCCTATTAGT

>h48
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTATCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCGTTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h49
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATACTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTACTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h50
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCAGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCCCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA

AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h51
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCTGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCCCTCTTTTAAGGATTAAGATTTAAGTCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGGTCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGATCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTTATAGCAGCACTATT
TACCTCCCTATTAGT

>h52
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTATCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCCCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h53
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATACTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAATTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATGTGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h54
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAACCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC

CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h55
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCGGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCCCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h56
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATACTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGATGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAGTAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCCCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATGTATATACTAATATGGGTCT
GCTCCTTACTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h57
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTT--ACCTGCTCACCCGCTCCTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTACTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h58
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC

TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTTAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h59
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCTGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCCCTCTTTTAAGGATTAAGATTTAAGTCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGAGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGATCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTTATAGCAGCACTATT
TACCTCCCTATTAGT

>h60
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCCCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGATAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h61
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCCCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGATAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h62
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCCTTTAAGGATTAAGATTTAAATCA

CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAGCACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h63
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATACTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGAGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h64
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGATGAACAGTTTATCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTCACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h65
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCGTTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGATAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h66
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGGGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG

AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCCCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h67
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCCTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h68
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAAAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGATAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h69
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTCCATGACCTTGATAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h70
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC

GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGATAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACGATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h71
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGAATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGATAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h72
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGATGAACAGTTTACCCTCCTTTATCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h73
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTTGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGATAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h74
CAACGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG

CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGATAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h75
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCTTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGGCCTTGGTAACTAAACACACCATTACTTAAATTAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAGACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTGTT
TACCTCCCTATTAGT

>h76
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCCAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h77
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAATATCGCC
CTCCGACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGAGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h78
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGATGAACAGTTTACCCTCCTTTGTCTAG

TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAGGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h79
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCATTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h80
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCCCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGCCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h81
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACGTCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGATAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h82
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA

TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCCCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCCCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h83
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAATATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGAGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h84
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTTTACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCCCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGAGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h85
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
TTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h86
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT

GACTCGTCCCACTAATATTAGGTGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h87
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGATGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACGTCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGAGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h88
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCACACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTTT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAAT

>h89
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATGTATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h90

CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGATAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h91
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGATAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAGACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h92
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTGTCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCCCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCACC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h93
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCCCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCACC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h94
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCTGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCCCTCTTTTAAGGATTAAGATTTAAGTCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCACC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGATCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTTATAGCAGCACTATT
TACCTCCCTATTAGT

>h95
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACGCCCAAGCCACTAATAGTTCATGACCTTGATAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h96
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTGTCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h97
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATAGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h98
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTACTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAGCACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h99
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCTGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAAAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCCCTCTTTTAAGGATTAAGATTTAAGTCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGATCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTTATAGCAGCACTATT
TACCTCCCTATTAGT

>h100
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTATCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAGCGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h101
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGAATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCCCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT

TACCTCCCTATTAGT

>h102
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTCCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAGTGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h103
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAAAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h104
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCTAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCCCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h105
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGATGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCCCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT

AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h106
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCCTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h107
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCGTCCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACGTCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h108
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGATGAACAGTTTACCCTCCTTTATCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCACTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCACACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCGAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h109
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATGCGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCCCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT

GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h110
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCACACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGATAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h111
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTGCTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h112
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTTATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h113
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATACTTACCACCCAACCAGAATAACTTACTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA

AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATGTGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h114
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
GGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h115
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACCAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h116
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAGCTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h117
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC

CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAAGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h118
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGGTTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h119
CAATGTAATTGTAACAGCCCATGCTTTTGTCATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGATAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h120
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAGCACCCAAGCCACTAATAGTTCATGACCTTGATAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h121
CAATGTAATTGTAACAGCCCATGCTTTTGTCATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC

TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAGTAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h122
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAAAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h123
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTGCCACCCAAGCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGGTGATCATACAAACCAAACTACTCCATTAGCACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h124
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGGCCTTGGTAACTAAACACACCATTACTTAAATTAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAGACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTGTT
TACCTCCCTATTAGT

>h125
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA

CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATTGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h126
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAGGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h127
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTGGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCGTCCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAGGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h128
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTTGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATCAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h129
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG

AGTAATGCTTACCACCCGACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAATATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h130
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCTCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCACTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGCCCTAAAACTAGGGGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTATTAGT

>h131
CAATGTAATTGTAACAGCCCATGCTTTTGTGATAATTTTCTTTCTAGTCATGCCCCTAATAATTGGCGGATTCGGTAACT
GACTCGTCCCACTAATATTAGGCGCACCAGATATGGCTTTTCCTCGACTCAACAACATAAGATTTTGACTACTTCCCCCA
TCCCTAACTCTTCTCCTAACATCAGGAGCCGTAGAAAGCGGTGCAGGGACAGGGTGAACAGTTTACCCTCCTTTGTCTAG
TAACCTTGCCCATGCAGGCGCATCCGTAGACCTTGCAATTTTCTCCCTACACTTAGCCGGAATCTCATCAATCCTAGGAG
CTGTAAACTTTATCACCACAGTAATAAACATACGAGTAAAAGCACAGCCCCTAGAACGGATACCTTTATTCGTCTGATCC
GTAAAAATTACCGCCATTCTACTTCTCCTATCCCTCCCTGTCCTTGCGGGAGCCATTACCGTCGATAATTTCTCAGAATG
AGTAATGCTTACCACCCAACCAGAATAACTTATTCTTTT-ACCTGCGCACCCGCTCTTTTAAGGATTAAGATTTAAATCA
CGTGTACCGCACATAAAGGCCACCAATCCTCCATGCCACAACTTGCACCAATTAATTGATTACTCCTATTCATCTTATTC
TGACTTACCGTAACAATCGTAGCCGTCATCATTTGATGATCATACAAACCAAACTACTCCATTAACACCCCAACATCGCC
CTCCAACAACACCCAAGCCACTAATAGTTCATGACCTTGGTAACTAAACACACCATTACTTAAATCAAACCACACCCAAA
AACAATGATAGCAGATATCTTTTCATCCTTCGACGACCATAATTCAGTCTTCATATCAATATATATACTAATATGGGTCT
GCTCCTTGCTAGTCTTACTCATTTTCAACATATCAATCTGAACAAATACAAACCGTTTTAGCTACCTACTAAACGTACCT
AAAAGGGTCATCTCATCCCAAGTCACACGATCCTTTGGCCTAAAACTAGGAGGATTTACAAATTTCATAGCAGCACTATT
TACCTCCCTGTTAGT

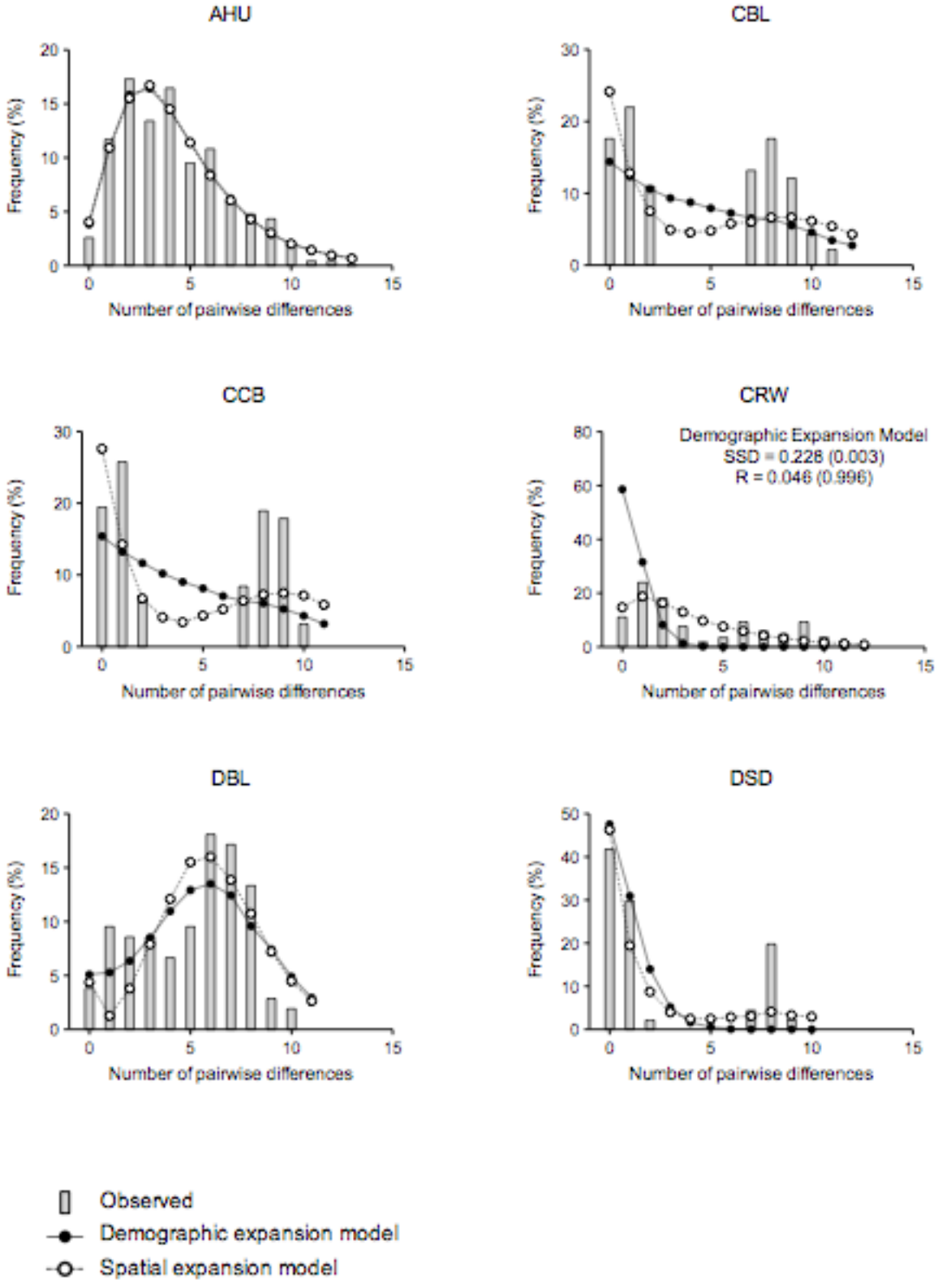
APPENDIX 3: MITOCHONDRIAL HAPLOTYPE FREQUENCIES

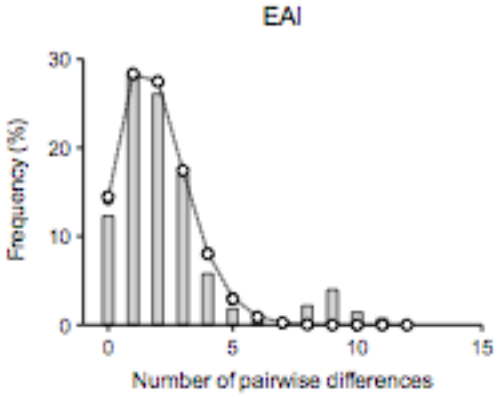
Hap ID	AHU	CBL	CCB	CRW	DBL	DSD	EAI	GLN	GOB	IHM	JCH	MAT	MTB	NPT	OCH	OLB	OPT	PHD	SPB	STR	TCL	TIM	TSK	WLG	WST	TOT
h0	1																				1					2
h1	1																									1
h2	1																									1
h3	2			3	1		5	2	1			1					2		2		1		1	1		22
h4	1																									1
h5	2							2		1											1			1		7
h6	1																									1
h7	1																									1
h8	3	3	4	3	2	2	6		8	2	5	4	7	5	2	1	2	3	3	3	2	4	6	1	6	87
h9	1							2																		3
h10	2	4	7	5	3	9	4	3	4	3	3	5	6	4		2	5	3	1	5	6	7	5	4	4	104
h11	1									1																2
h12	1		2	2						2							2	1	1	1				1		13
h13	1																									1
h14	1																		1							2
h15	1																									1
h16	1			2														1								4
h17		4	4			2			3		5	1		5	5	4		3		4	2	1	2		2	47
h18		2	3			1	3		1		4	1	4	2	3	4		1		1	1	5	3	2	2	43
h19		1																								1
h20				2				1									1				1					5
h21				1																						1
h22				1																						1
h23					1																					1
h24					1																					1
h25					1																			1		2
h26					1																					1
h27					1																					1
h28					1																					1
h29					1																					1
h30					1																					1
h31					1																					1
h32							1																			1
h33							1														1					2
h34							1																			1
h35							1																			1
h36							1				1	1	1	1								1				6
h37							1	1								1										3
h38								1		1																2
h39								1																		1
h40								1									1									2
h41								1																		1
h42								2				1														3
h43								1																		1
h44								1																		1
h45								1																		1
h46									1																	1
h47									1																	1

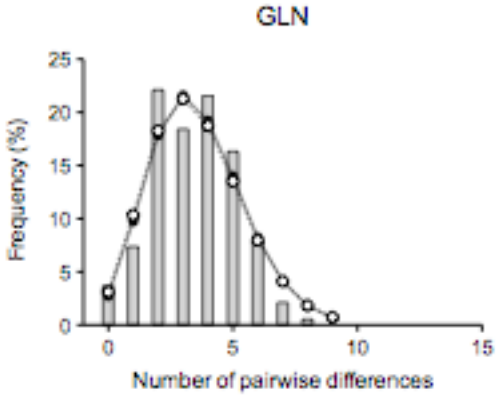
Hap ID	AHU	CBL	CCB	CRW	DBL	DSD	EAI	GLN	GOB	IHM	JCH	MAT	MTB	NPT	OCH	OLB	OPT	PHD	SPB	STR	TCL	TIM	TSK	WLG	WST	TOT
h48										1																1
h49										1																1
h50										1																1
h51										1																1
h52										1																1
h53										1																1
h54										1																1
h55										1							1									2
h56										1																1
h57										1																1
h58											1															1
h59											1															1
h60												1														1
h61												1														1
h62												1														1
h63												1														1
h64												1														1
h65												1							1							2
h66												1														1
h67													1													1
h68													1													1
h69													1													1
h70													1													1
h71														1												1
h72														1												1
h73														1											1	2
h74															1											1
h75															1											1
h76															1											1
h77																2										2
h78																1										1
h79																1										1
h80																1										1
h81																1										1
h82																1										1
h83																1										1
h84																	1									1
h85																	1									1
h86																	1									1
h87																	1									1
h88																	1									1
h89																	1									1
h90																		1								1
h91																		1								1
h92																		1								1
h93																		1								1
h94																		1								1
h95																		1								1

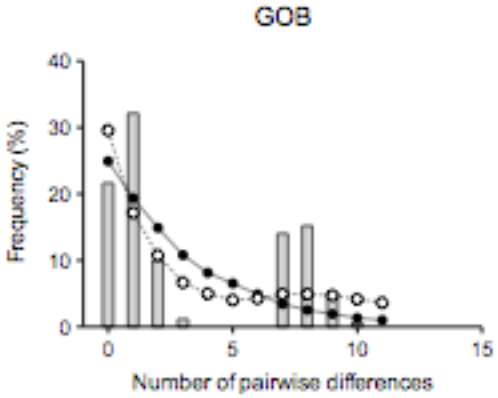
Hap ID	AHU	CBL	CCB	CRW	DBL	DSD	EAI	GLN	GOB	IHM	JCH	MAT	MTB	NPT	OCH	OLB	OPT	PHD	SPB	STR	TCL	TIM	TSK	WLG	WST	TOT
h96																		1								1
h97																		1								1
h98																			1							1
h99																			1							1
h100																			1							1
h101																			1							1
h102																			1							1
h103																			1							1
h104																			1							1
h105																			1							1
h106																			1							1
h107																			1							1
h108																			1							1
h109																			1							1
h110																				1						1
h111																					1					1
h112																					1					1
h113																					1					1
h114																					1					1
h115																						1				1
h116																						1				1
h117																						1				1
h118																						1				1
h119																						1				1
h120																							1			1
h121																							1			1
h122																							1			1
h123																								1		1
h124																								1		1
h125																								1		1
h126																								1		1
h127																								1		1
h128																								1		1
h129																								1		1
h130																								1		1
h131																								1		1
Total	22	14	20	19	15	14	24	20	19	20	20	21	22	20	13	20	20	20	21	15	20	23	20	20	15	477

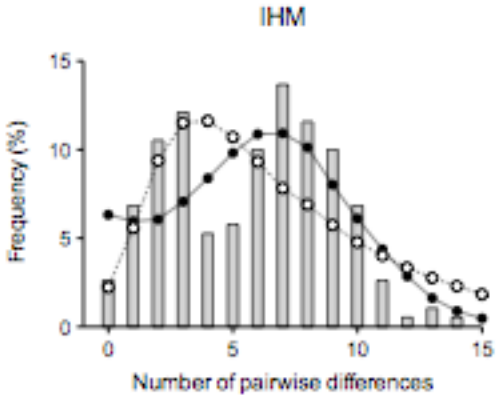
APPENDIX 4: MISMATCH DISTRIBUTIONS

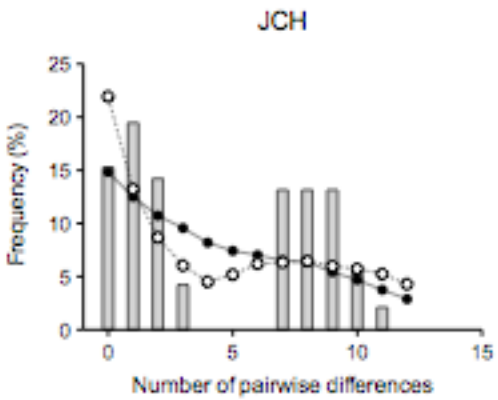


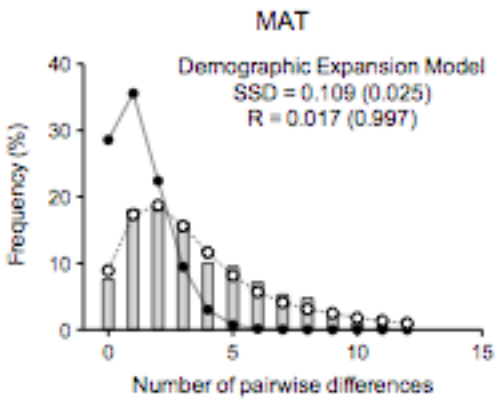

EAI
Frequency (%)
Number of pairwise differences

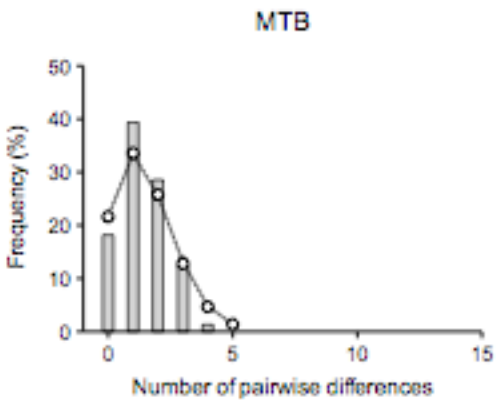

GLN
Frequency (%)
Number of pairwise differences

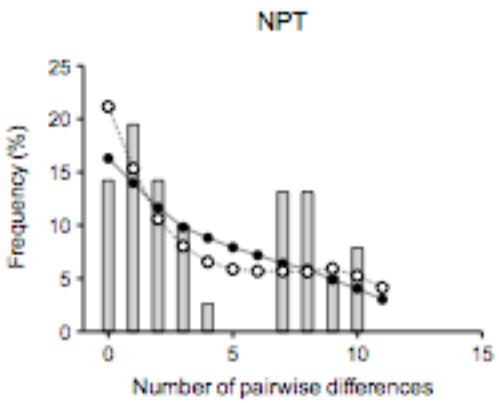

GOB
Frequency (%)
Number of pairwise differences

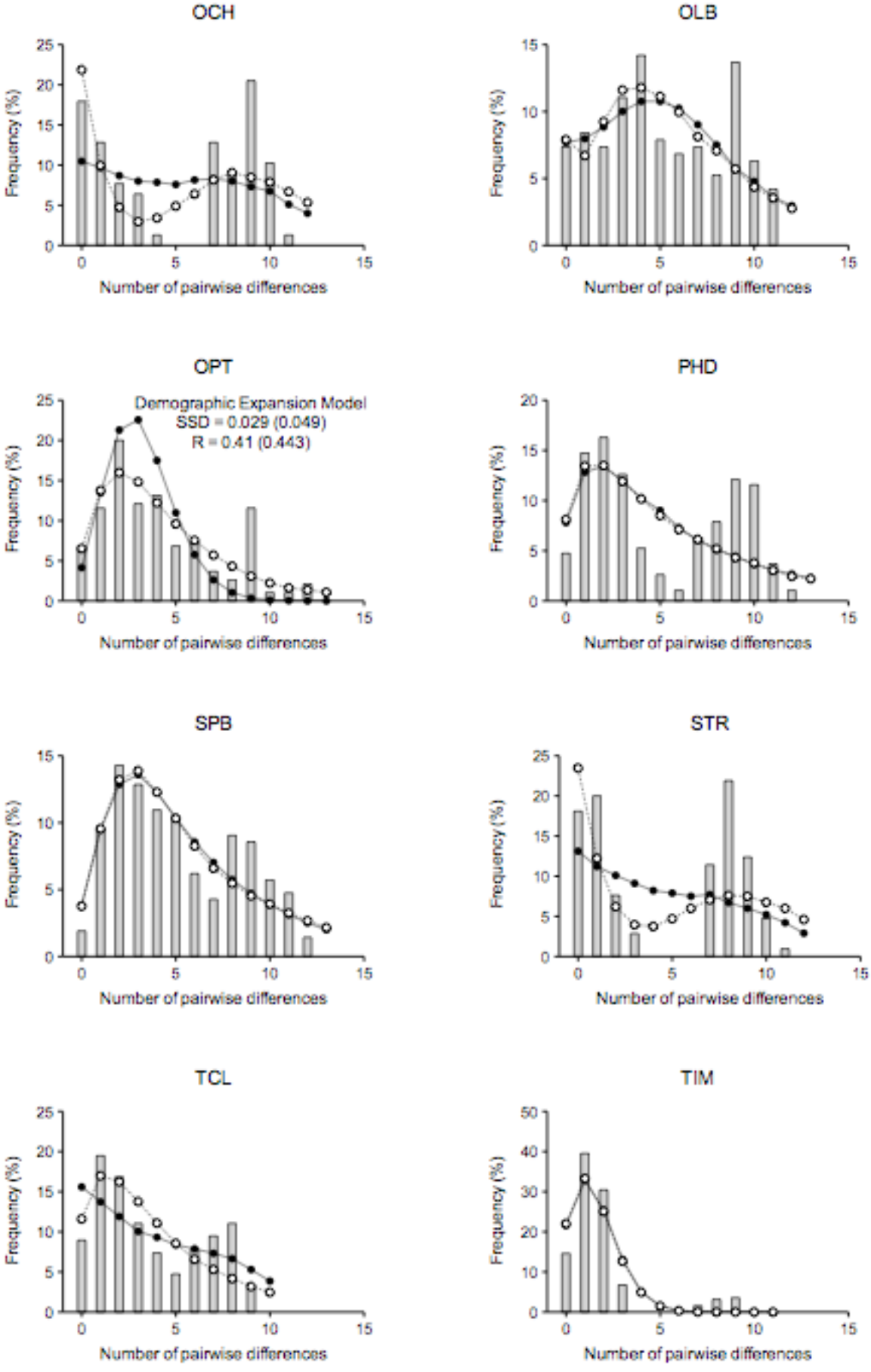

IHM
Frequency (%)
Number of pairwise differences


JCH
Frequency (%)
Number of pairwise differences


MAT
Demographic Expansion Model
SSD = 0.109 (0.025)
R = 0.017 (0.997)
Frequency (%)
Number of pairwise differences


MTB
Frequency (%)
Number of pairwise differences


NPT
Frequency (%)
Number of pairwise differences


OCH
OLB
OPT
Demographic Expansion Model
SSD = 0.029 (0.049)
R = 0.41 (0.443)
PHD
SPB
STR
TCL
TIM
Frequency (%)
Number of pairwise differences

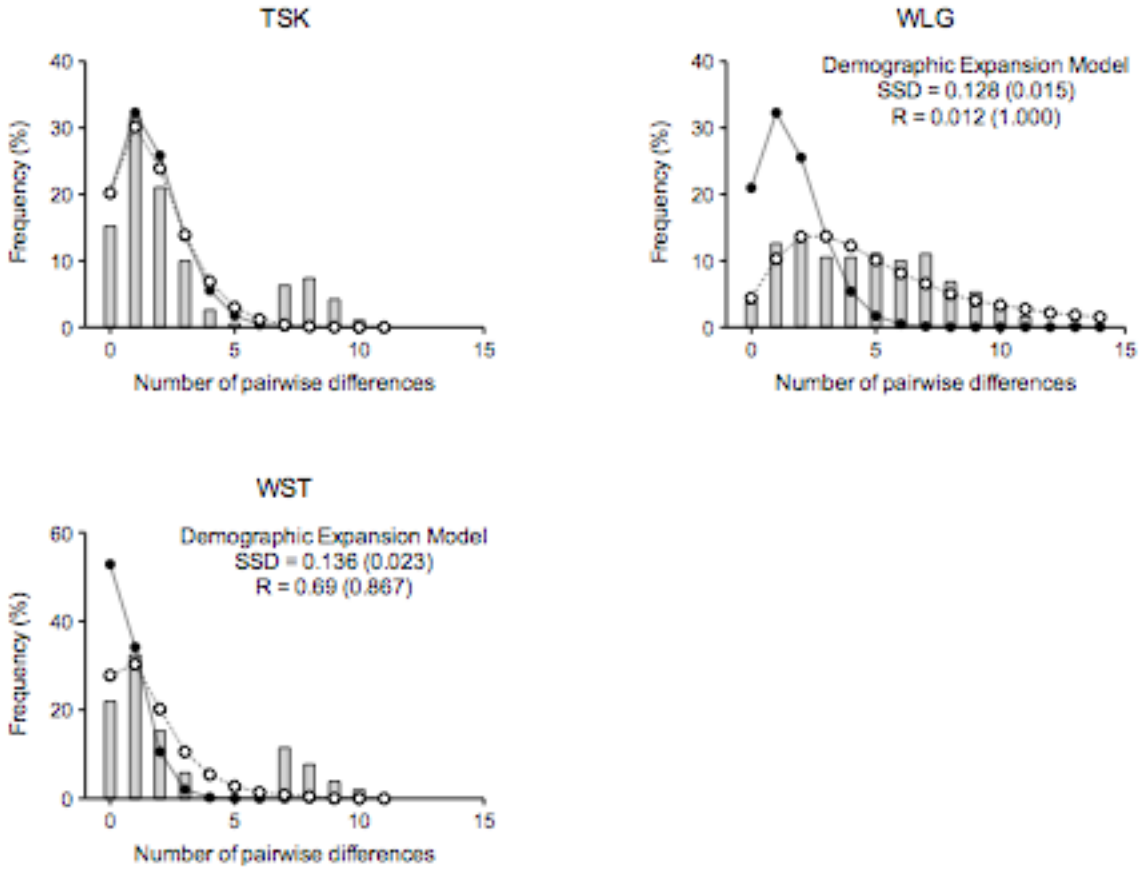
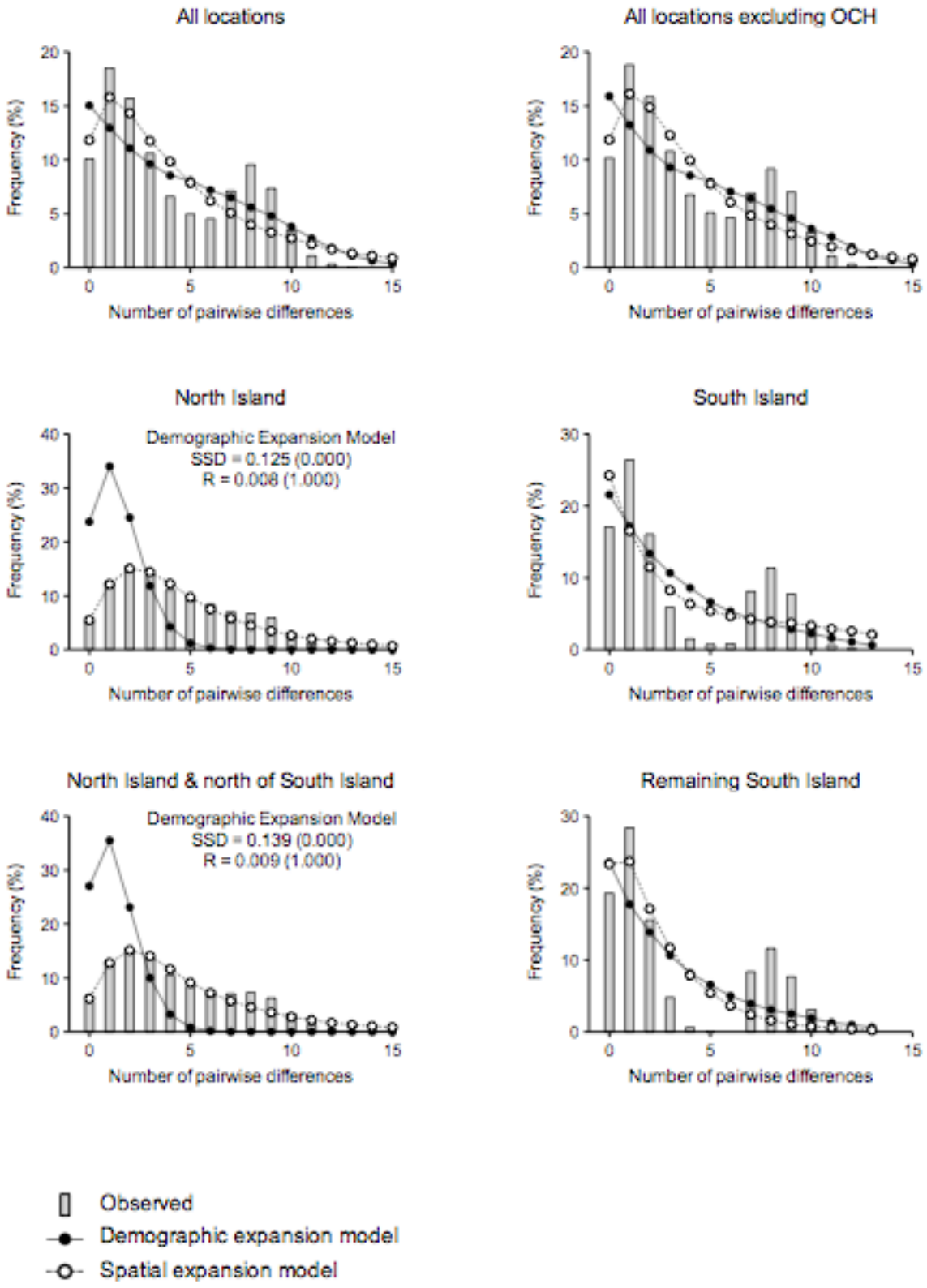


Figure A5.1: Mismatch distributions for sampling locations. Frequencies of observed pairwise differences (gray bars) were compared to two different models of population expansion: pure demographic expansion (closed circles and solid lines) and spatial population expansion (open circles and dashed lines). Model fit was tested using 1000 bootstraps in Arlequin 3.1 (Excoffier et al. 2005). Only significant sum of squared deviations (SSD) and p-values were reported for a given model. Raggedness indices (R, Harpending et al. 1993; Harpending 1994) were also reported for models with significant SSD.


All locations
Frequency (%)
Number of pairwise differences
All locations excluding OCH
Frequency (%)
Number of pairwise differences
North Island
Demographic Expansion Model
SSD = 0.125 (0.000)
R = 0.008 (1.000)
Frequency (%)
Number of pairwise differences
South Island
Frequency (%)
Number of pairwise differences
North Island & north of South Island
Demographic Expansion Model
SSD = 0.139 (0.000)
R = 0.009 (1.000)
Frequency (%)
Number of pairwise differences
Remaining South Island
Frequency (%)
Number of pairwise differences
Observed
Demographic expansion model
Spatial expansion model

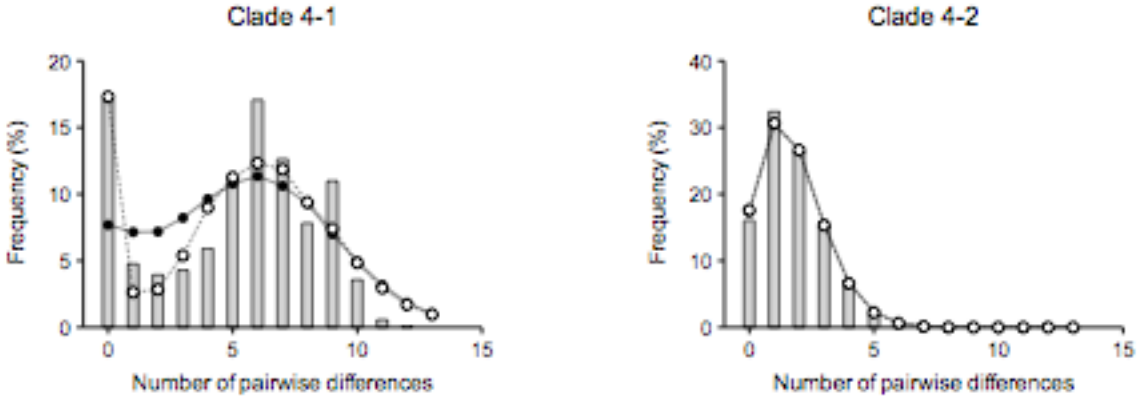


Figure A5.2: Mismatch distributions for proposed in Table 2.2 and clades 4-1 and 4-1 in Figure 2.11. Frequencies of observed pairwise differences (gray bars) were compared to two different models of population expansion: pure demographic expansion (closed circles and solid lines) and spatial population expansion (open circles and dashed lines). Model fit was tested using 1000 bootstraps in Arlequin 3.1 (Excoffier et al. 2005). Only significant sum of squared deviations (SSD) and p-values were reported for a given model. Raggedness indices (R, Harpending et al. 1993; Harpending 1994) were also reported for models with significant SSD.

APPENDIX 5: ABALONE MICROSATELLITES

Table A5.1: Abalone microsatellites. Listed are the species in which the microsatellite locus was amplified, the aim of the study (D, microsatellite development; A, application in a population genetics; C, species-specific microsatellite conservation; if only an A is listed then the locus was isolated in another study), the locus name and its GenBank accession number (ACCN), repeat array (N, pure; n, interrupted), the number of individuals genotyped (N), the size or size range of the amplified fragment, the number of alleles (N_A), the observed heterozygosity (H_O), and the expected heterozygosity (H_E; *significant Hardy-Weinberg disequilibrium, ^{NA} Hardy-Weinberg equilibrium test not available or confusing). This table incorporates most of the abalone microsatellite studies but is not exhaustive. Although some articles report data as the mean across loci and these articles are included in the table, data such as the number of alleles, observed heterozygosity, and expected heterozygosity are only included if could be easily discerned from the literature. Deviations from HWE were taken at $p < 0.05$ OR $p < 0.01$, according to the original article.

Species	Aim	Locus ACCN	Motif	N	Size (bp)	N_A	H_O	H_E	Reference
Eastern Pacific									
H. kamtschatkana									
	D	Hka3 AY013574	$(GTA)_N(GAGT)_N$	442	229–314	51	0.44	0.96*	(Miller et al. 2001)
	D	Hka6 AY013580	$(TGG)_N(GGC)_N$ $(GGT)_N(AGG)_N$ $(AGGG)_N(AGG)_N$	621	106–188	20	0.47	0.72*	(Miller et al. 2001)
	D	Hka12 AY013572	$(CA)_n$	1162	184–363	63	0.89	0.92	(Miller et al. 2001)
	D	Hka28 AY013573	$(CA)_N$	567	187–249	30	0.55	0.94*	(Miller et al. 2001)
	D	Hka37 AY013575	$(TGG)_N(GGT)_N$ $(GGA)_N(GGGA)_n$	424	243–305	21	0.65	0.68	(Miller et al. 2001)
	D	Hka40 AY013576	$(CA)_N$	643	110–180	24	0.85	0.91	(Miller et al. 2001)
	D	Hka43 AY013577	$(GACA)_N$	567	179–255	20	0.87	0.88	(Miller et al. 2001)
	D	Hka48 AY013578	$(CA)_n$	438	120–220	52	0.72	0.95*	(Miller et al. 2001)
	D	Hka56 AY013579	$(CA)_N$	659	97–148	26	0.85	0.92	(Miller et al. 2001)
	D	Hka65 AY013581	$(CA)_N(CA)_N$	637	100–200	37	0.83	0.93*	(Miller et al. 2001)
	D	Hka80 AY013582	$(CA)_N$	441	88–144	27	0.89	0.92	(Miller et al. 2001)
	D	Hka85 AY013583	$(ACC)_N(ACT)_N$ $(ACC)_N(ACT)_N$	501	130–340	23	0.41	0.90*	(Miller et al. 2001)
	A	Hka12		3345	171-377	82	0.89	0.92^{NA}	(Withler et al. 2001)
	A	Hka28		3345	183-271	37	0.57	0.94^{NA}	(Withler et al. 2001)
	A	Hka40		3345	112-210	37	0.85	0.91^{NA}	(Withler et al. 2001)
	A	Hka43		3345	163-263	24	0.87	0.88^{NA}	(Withler et al. 2001)
	A	Hka48		3345	93-250	68	0.71	0.97^{NA}	(Withler et al. 2001)
	A	Hka56		3345	93-164	35	0.86	0.92^{NA}	(Withler et al. 2001)
	A	Hka65		3345	115-250	58	0.87	0.95^{NA}	(Withler et al. 2001)
	A	Hka85		3345	122-390	49	0.49	0.89^{NA}	(Withler et al. 2001)
H. fulgens									
	D	Hful240 AY952206	$(TC)_NCC$	25	154	1	–	–	(Cruz et al. 2005)

	D	Hful333 AY952207	$(TC)_N$	25	172	1	–	–	(Cruz et al. 2005)
	D	Hful369 AY952208	$(GA)_N$	22	268–325	20	0.73	0.93	(Cruz et al. 2005)
	D	Hful442 AY952209	$(TC)_N(AC)_N$	25	154–159	3	1	0.62*	(Cruz et al. 2005)
	D	Hful547 AY952210	$(AT)_N$TATTA	13	215–237	13	1	0.88*	(Cruz et al. 2005)
	D	Hful603 AY952211	$(TG)_N$	25	186–204	5	0.52	0.60	(Cruz et al. 2005)
	D	Hful910 AY952212	$(TG)(TG)_n(GT)_N$	24	142–156	5	0.67	0.57	(Cruz et al. 2005)
	D	Hful1136 AY952213	$(C)_N$	21	209–233	10	0.52	0.73*	(Cruz et al. 2005)
	A	Hka28		209		–	–	–	(Gutiérrez-Gonzalez and Perez-Enriquez 2005)
	A	Hka56		209		–	–	–	(Gutiérrez-Gonzalez and Perez-Enriquez 2005)
	A	Hka28		460		52	–	–	(Gutiérrez-Gonzalez et al. 2007)
	A	Hka56		432		29	–	–	(Gutiérrez-Gonzalez et al. 2007)
	A	Hful603		427		5	–	–	(Gutiérrez-Gonzalez et al. 2007)
	D, A	Hful260		423		4	–	–	(Gutiérrez-Gonzalez et al. 2007)
H. corrugata									
	D	Hco6 EF694951	$(CA)_N$	42	223–271	17	0.405	0.830*	(Díaz-Viloria et al. 2008)
	D	Hco15 EF694952	$(TCAC)_N(CT)_N(TCAC)_N$	49	195–235	10	0.510	0.516	(Díaz-Viloria et al. 2008)
	D	Hco16 EF694953	$(TTG)_N$	49	204–246	8	0.653	0.601	(Díaz-Viloria et al. 2008)
	D	Hco19 EF694954	$(TG)_N(TGCG)_N$	48	151–191	17	0.646	0.860	(Díaz-Viloria et al. 2008)
	D	Hco22 EF694955	$(CTCG)_N$	49	211–225	6	0.694	0.754	(Díaz-Viloria et al. 2008)
	D	Hco23	$(TGAG)_N$	48	222–233	4	0.104	0.213*	(Díaz-Viloria et al. 2008)

		EF694956							
	D	Hco43 EF694957	$(CA)_N(CA)_N(CA)_N$	42	192–197	2	0.143	0.433*	(Díaz-Viloria et al. 2008)
	D	Hco47 EF694958	$(AC)_N(TCAC)_N$ $TCAT(TCAC)_N$	47	238–248	6	0.617	0.686	(Díaz-Viloria et al. 2008)
	D	Hco97 EF694959	$(CT)_n$	49	177–211	7	0.551	0.497	(Díaz-Viloria et al. 2008)
	D	Hco194 EF694960	$(TA)_N$	49	196–198	2	0.224	0.262	(Díaz-Viloria et al. 2008)
	D	Hka13 EU090247	$(GTA)_N(GAGT)_N$	49	193–332	55	0.939	0.982	(Díaz-Viloria et al. 2008)
	D	Hka56 EU090246	$(CA)_N$	48	237–254	4	0.500	0.515	(Díaz-Viloria et al. 2008)
H. sorenseni									
	A	Hka3		19		–	0.74	0.56	(Gruenthal and Burton 2005)
	A	Hka28		19		–	0.53	0.56	(Gruenthal and Burton 2005)
	A	Hka40		18		–	0.78	0.73	(Gruenthal and Burton 2005)
	A	Hka56		19		–	0.68	0.65	(Gruenthal and Burton 2005)
	A	Hka80		19		–	0.79	0.76	(Gruenthal and Burton 2005)
H. rufescens									
	A	Hka3		445		75	0.90	0.96	(Gruenthal et al. 2007)
	A	Hka28		429		29	0.86	0.88	(Gruenthal et al. 2007)
	A	Hka40		448		28	0.91	0.91	(Gruenthal et al. 2007)
	A	Hka56		401		26	0.42	0.80*	(Gruenthal et al. 2007)
	A	Hka80		396		36	0.73	0.83*	(Gruenthal et al. 2007)
	D	Hruf200		74	97–149	21	0.47	0.76	(Kirby et al. 1998)
H. cracherodii									
	A	Hka28		579		33	0.88	0.95	(Gruenthal and Burton 2008)
	A	Hka40		437		51	0.21	0.94	(Gruenthal and Burton 2008)
	A	Hka56		522		48	0.34	0.94	(Gruenthal and Burton 2008)
	A	Hka80		520		51	0.40	0.96	(Gruenthal and Burton 2008)
North Pacific									
H. discus discus									
	D	Hdd6C AB025367	$(GACT)_N(CTCA)_N(CA)_n$	20	219–247	7	0.80	0.81	(Sekino and Hara 2001)
	D	Hdd108C AB025384	$(CA)_N$	20	170–186	5	0.55	0.57	(Sekino and Hara 2001)
	D	Hdd114B AB025387	$(CA)_n(CGCA)_N(CA)_N$	36	216–250	10	0.78	0.89*	(Sekino and Hara 2001)
	D	Hdd115B	$(CA)_N(CA)_N$	24	243–247	3	0.17	0.20	(Sekino and Hara 2001)

		AB025388							
	D	Hdd229 AB47107	$(TCA)_N(AT)_N(TA)_NX_N(GA)_N(CA)_N(CTCA)_NX_N(CA)_N$	24	190–220	8	0.67	0.81*	(Sekino and Hara 2001)
H. discus hannai									
	D, A	Hd527 AB178064	$(CTCA)_N$				–	–	(Hara and Sekino 2005)
	D, A	Hd535 AB178065	$(CTCA)_N$				–	–	(Hara and Sekino 2005)
	D, A	Hd601 AB178069	$(CGCA)_N(CA)_N$				–	–	(Hara and Sekino 2005)
	D, A	Hd604 AB178070	$(AAT)_N$				–	–	(Hara and Sekino 2005)
	D, A	Hd680 AB178073	$(CTCA)_N$				–	–	(Hara and Sekino 2005)
	D, A	Hd715 AB178074	$(CTCA)_N$				–	–	(Hara and Sekino 2005)
	D, A	Hd724 AB178076	$(TC)_NT_NA(CT)_N$				–	–	(Hara and Sekino 2005)
	D, A	Hd731 AB178077	$(ATG)_N$				–	–	(Hara and Sekino 2005)
	D	co89 DQ992519	$(TG)_N$	31	331–395	5	0.452	0.647	(Sun et al. 2007)
	D	ca324 DQ992520	$(AT)_n$	31	195–295	12	0.613	0.878*	(Sun et al. 2007)
	D	co221 DQ992521	$(CAA)_N$	29	256–303	2	0.469	0.424	(Sun et al. 2007)
	D	ca481 DQ992522	$(AGC)_N$	32	162–203	4	0.194	0.182	(Sun et al. 2007)
	D	ca465 DQ992523	$(AAGT)_N$	32	184–276	12	0.40	0.871*	(Sun et al. 2007)
	D	co11 DQ992524	$(ACTC)_N$	32	296–332	6	0.25	0.405*	(Sun et al. 2007)
	D	ca333 DQ992525	$(TCTG)_N$	32	247–284	9	0.484	0.809*	(Sun et al. 2007)
	D	ca235 DQ992526	$(AAG)_N$	31	276–326	2	0.656	0.496	(Sun et al. 2007)
	D	aa102 DQ992527	$(TA)_n$	31	277–293	2	0.524	0.396	(Sun et al. 2007)
	D	co220 DQ992528	$(AG)_N$	31	239–257	3	0.774	0.556*	(Sun et al. 2007)

	D	ba139 DQ992529	$(AGAC)_N$	31	156–176	3	0.276	0.246	(Sun et al. 2007)
	D	ca541 DQ992530	$(TC)_N$	32	227–231	2	0.25	0.437*	(Sun et al. 2007)
	D	co4 DQ992531	$(AG)_NAT(TG)_N$	32	310–322	2	0.969	0.612*	(Sun et al. 2007)
	D	ca955 DQ992532	$(AT)_N$	30	407–409	2	0.594	0.504	(Sun et al. 2007)
	D	co119 DQ992533	$(TA)_n$	32	268–274	2	0.094	0.091	(Sun et al. 2007)
	D	KHdh6 AY948316	$(CA)_n$	30	238–246	3	0.50	0.42	(An and Han 2006)
	D	KHdh28 AY948312	$(AC)_N$	30	132–180	17	0.87	0.93	(An and Han 2006)
	D	KHdh43 AY948314	$(GT)_N$	30	172–280	28	0.87	0.92	(An and Han 2006)
	D	KHdh44 AY948317	$(TG)_n$	30	136–202	24	0.90	0.96	(An and Han 2006)
	D	KHdh46 AY948320	$(AC)_n$	30	226–248	10	0.93	0.95*	(An and Han 2006)
	D	KHdh47 AY948321	$(AC)_n$	30	226–250	11	1	0.83	(An and Han 2006)
	D	KHdh50 AY948318	$(CA)_n$	30	130–150	5	0.67	0.80*	(An and Han 2006)
	D	KHdh53 AY948319	$(TTC)_N(TTG)_N$	30	180–286	4	0.97	0.60*	(An and Han 2006)
	D	KHdh55 AY948325	$(AC)_N(AT)_N(AC)_N$	30	162–176	6	0.17	0.52*	(An and Han 2006)
	D	KHdh57 AY948326	$(ACGC)_N(AC)_n$	30	236–266	10	1	0.71*	(An and Han 2006)
	D	KHdh59 AY948327	$(AAC)_n$	30	238–264	11	0.7	0.85	(An and Han 2006)
	D	KHdh64 AY948328	$(AC)_N$	30	242–252	6	0.83	0.86*	(An and Han 2006)
	D	KHdh66 AY948329	$(GAGT)_n$	30	98–170	18	0.43	0.70	(An and Han 2006)
	D	KHdh76 AY948331	$(GT)_n(GCGT)_n$	30	190–254	25	0.8	0.92	(An and Han 2006)
	D	KHdh80 AY948332	$(TCAC)_n(TG)_n$	30	206–322	21	0.93	0.96	(An and Han 2006)

	D	KHdh86 AY948315	$(AC)_n$	30	120–130	5	0.5	0.64	(An and Han 2006)
	D	KHdh89 AY948336	$(AC)_N$	30	140–204	19	0.97	0.9	(An and Han 2006)
	D	Afa002 AB177903	$(CA)_NA(AC)_N$	97	163		–	–	(Sekino et al. 2005)
	D	Afa005 AB177904	$(AC)_N$	97	192		–	–	(Sekino et al. 2005)
	D	Afa014 AB177905	$(CGCA)_NCA(CACG)_N(CA)_NN_N(AC)_N$	97	196		–	–	(Sekino et al. 2005)
	D	Afa017 AB177906	$(AC)_NN_N(ACGC)_NN_N(ACGC)_N$	97	201		–	–	(Sekino et al. 2005)
	D	Afa020 AB177907	$(AC)_N(TAACACACACAC)_NTA(AC)_NN_N(AC)_N$	97	219		–	–	(Sekino et al. 2005)
	D	Awb002 AB177908	$(CA)_N$	97	214		–	–	(Sekino et al. 2005)
	D	Awb003 AB177909	$(TGAG)_N$	97	257		–	–	(Sekino et al. 2005)
	D, A	Awb009 AB177910	$(TGAG)_N$	97	235		–	–	(Sekino et al. 2005)
	D	Awb016 AB177911	$(GA)_NN_N(GAGAGAAA)_N(GA)_NA(AG)_NT(GA)_NA_N(AG)_NT(GA)_NA(AG)_NT(GA)_NN_N(GA)_N$	97	196		–	–	(Sekino et al. 2005)
	D, A	Awb017 AB177912	$(CA)_N$	97	215		–	–	(Sekino et al. 2005)
	D	Awb019 AB177913	$(GAGT)_N$	97	250		–	–	(Sekino et al. 2005)
	D, A	Awb022 AB177914	$(TG)_N$	97	201		–	–	(Sekino et al. 2005)
	D, A	Awb026 AB177915	$(ACCCACAC)_N$	97	167		–	–	(Sekino et al. 2005)
	D	Awb027 AB177916	$(CG)_N(CA)_N(CG)_N(CGCA)_NC(AT)_N\ (AC)_N\ N_N(CA)_NN_N\ (AC)_N\ (TC)_N(AC)_NN_N(CG)_NCT(CA)_NCG(CT)_NA(TC)_N(AC)_N$	97	275		–	–	(Sekino et al. 2005)
	D	Awb028 AB177917	$(AC)_N$	97	162		–	–	(Sekino et al. 2005)
	D, A	Awb033 AB177918	$(AAT)_N$	97	175		–	–	(Sekino et al. 2005)

	D	Awb035 AB177919	$(TA)_NAT(CA)_N$	97	209		–	–	(Sekino et al. 2005)
	D	Awb036 AB177920	$(CA)_N(CG)_N(CA)_N$	97	187		–	–	(Sekino et al. 2005)
	D	Awb037 AB177921	$(CACG)_N(CA)_N$	97	191		–	–	(Sekino et al. 2005)
	D	Awb038 AB177922	$(CA)_N(CACG)_NCG(CA)_N(CACG)_N(CA)_N$ $(CACG)_NN_N(CG)_N$ $(CA)_NN_N(AC)_NG(CA)_N$	97	189		–	–	(Sekino et al. 2005)
	D, A	Awb039 AB177923	$(ATT)_N$	97	174		–	–	(Sekino et al. 2005)
	D	Awb041 AB177924	$(ATG)_N$	97	201		–	–	(Sekino et al. 2005)
	D	Awb042 AB177925	$(AC)_N$	97	199		–	–	(Sekino et al. 2005)
	D	Awb044 AB177926	$(AC)_N(GC)_NN_N(CT)_NN_N(CA)_N$	97	160		–	–	(Sekino et al. 2005)
	D	Awb052 AB177927	$(CA)_N$	97	228		–	–	(Sekino et al. 2005)
	D	Awb058 AB177928	$(TA)_n$	97	200		–	–	(Sekino et al. 2005)
	D, A	Awb062 AB177929	$(ATT)_N$	97	235		–	–	(Sekino et al. 2005)
	D	Awb063 AB177930	$(AC)_N(GC)_NA(CGCA)_N(CA)_N(CACG)_N(CA)_N$	97	210		–	–	(Sekino et al. 2005)
	D	Awb068 AB177931	$(CT)_N$	97	217		–	–	(Sekino et al. 2005)
	D	Awb071 AB177932	$(CT)_N$	97	184		–	–	(Sekino et al. 2005)
	D	Awb074 AB177933	$(CACG)_NN_N(AC)_N(GC)_N(AC)_N$	97	194		–	–	(Sekino et al. 2005)
	D	Awb076 AB177934	$(TA)_N$	97	198		–	–	(Sekino et al. 2005)
	D, A	Awb079 AB177935	$(TGAG)_N(GAGT)_N$	97	199		–	–	(Sekino et al. 2005)
	D, A	Awb083 AB177936	$(ATC)_N$	97	238		–	–	(Sekino et al. 2005)
	D	Awb089 AB177937	$(TC)_N$	97	203		–	–	(Sekino et al. 2005)

	D	Awb091 AB177938	$(TC)_N$	97	186		–	–	(Sekino et al. 2005)
	D	Awb098 AB177939	$(AC)_N$	97	184		–	–	(Sekino et al. 2005)
	D	Awb101 AB177940	$(AG)_N$	97	165		–	–	(Sekino et al. 2005)
	D	Hdh1321 AB084076	$(CGCA)_N(CA)_N$	30	272–362	20	0.97	0.92	(Li et al. 2002a)
	D	Hdh78 AB084077	$(CACCT)_n$	30	177–332	7	0.33	0.60*	(Li et al. 2002a)
	D	Hdh1761 AB084078	$(CA)_n…(CCACA)_N$	30	405–596	18	0.30	0.92*	(Li et al. 2002a)
	D	Hdh1457 AB084079	$(CGCCA)_N(CTCCA)_n$	30	481–601	12	0.33	0.71*	(Li et al. 2002a)
	A	Hdh1321		42			–	–	(Li et al. 2004)
	D, A	Hdh513 AB091483	$(GA)_N$	45			–	–	(Li et al. 2004)
	D, A	Hdh57 AB091479	$(CA)_N$	11			–	–	(Li et al. 2004)
	D, A	Hdh145 AB091480	$(CA)_N$	8			–	–	(Li et al. 2004)
	A	Hdd114B AB025387		29			–	–	(Li et al. 2004)
	A	Hdd108C AB025384		6			–	–	(Li et al. 2004)
	D	Afa029 AB239612	$(CA)_N(CACG)_N(CA)_n(CG)_n(CA)_N$		210				(Sekino et al. 2006)
	D	Afa034 AB239613	$(CACG)_N(CA)_NX_N(CGCACA)_N(CACG)_N(CA)_n$		206				(Sekino et al. 2006)
	D	Afa036 AB239616	$(AC)_N(ACGC)_N(AC)_N(ACGC)_N$		201				(Sekino et al. 2006)
	D	Afa037 AB239614	$(AC)_N$		217				(Sekino et al. 2006)
	D	Afa038 AB239615	$(AC)_N(ACGC)_NX_N(CTCA)_N(CA)_NT(AC)_n(AC)_N(CA)_N$		232				(Sekino et al. 2006)
	D	Afa039 AB239617	$(CA)_N(CACG)_N$		238				(Sekino et al. 2006)
	D	Afa040 AB239618	$(AC)_NX)_N(AC)_N(GC)_N$		217				(Sekino et al. 2006)

	D	Afa041 AB239619	$(CA)_NT(AC)_N(ACGC)_N(AC)_NX_N(ACGC)_NX_N(AC)_N$		237				(Sekino et al. 2006)
	D	Afa047 AB239620	$(CGCA)_N(CA)_N(CACG)_N(CA)_N$		222				(Sekino et al. 2006)
	D	Afa049 AB239621	$(CAGA)_N(CA)_N$		212				(Sekino et al. 2006)
	D	Afa050 AB239622	$(AC)_N$		216				(Sekino et al. 2006)
	D	Afa051 AB239623	$(CACG)_N(CA)_NT(AC)_N$		209				(Sekino et al. 2006)
	D	Afa054 AB239624	$(AC)_NX_N(CACG)_NX_N(CACG)_N(CA)_N$		160				(Sekino et al. 2006)
	D	Afa061 AB239625	$(CACACG)_N(CACACAGAC)_N(AC)_n(AG)_N\ X_N(AC)_N$		250				(Sekino et al. 2006)
	D	Afa066 AB239626	$(CA)_N$		216				(Sekino et al. 2006)
	D	Afa068 AB239627	$(CA)_NA(AC)_NX_N(CA)_N$		234				(Sekino et al. 2006)
	D	Afa071 AB239628	$(CA)_N$		198				(Sekino et al. 2006)
	D	Afa075 AB239629	$(CA)_NX_N(GCAC)_N$		169				(Sekino et al. 2006)
	D	Afa077 AB239630	$(GA)_N(CA)_N(AC)_NA_N(AC)_N(ACGC)_N(AC)_N$		224				(Sekino et al. 2006)
	D	Afa080 AB239631	$(CT)_N(CACT)_NCT(CA)_N$		231				(Sekino et al. 2006)
	D	Afa083 AB239632	$(CA)_NX_N(AC)_NX)_N(CACACGCA)_N(CA)_N$		223				(Sekino et al. 2006)
	D	Afa084 AB239633	$(CA)_N(CACGCA)_N(CA)_N(CACG)_N(CA)_N$		226				(Sekino et al. 2006)
	D	Afa093 AB239634	$(AC)_N(GC)_N$		191				(Sekino et al. 2006)
	D	Afa095 AB239635	$(AC)_NA_N(AC)_N(GCAC)_N(TC)_N(AC)_N(ACACGC)_N$		235				(Sekino et al. 2006)
	D	Afa097 AB239636	$(CA)_N$		186				(Sekino et al. 2006)
	D	Afa098 AB239637	$(CA)_N(CG)_N(CA)_N$		225				(Sekino et al. 2006)
	D	Afa099 AB239638	$(CA)_N(CACG)_N(CA)_N(CCACCACCAGCA)_N(CCA)_N$		224				(Sekino et al. 2006)

	D	Afa100 AB239639	$(AC)_N(AG)_N(GC)_N(AC)_N(AG)_N(GCAC)_N(GC)_N$		240				(Sekino et al. 2006)
	D	Afa101 AB239640	$(TC)_nA(CT)_nG(TC)_N$		237				(Sekino et al. 2006)
	D	Afa105 AB239641	$(CACG)_N(CG)_NX_N(CA)_N$		215				(Sekino et al. 2006)
	D	Afa107 AB23942	$(CACG)_N$		209				(Sekino et al. 2006)
	D	Afa109 AB239643	$(CTCA)_N$		182				(Sekino et al. 2006)
	D	Afa110 AB23964	$(ACTC)_N$		224				(Sekino et al. 2006)
	D	Afa115 AB239645	$(TCAC)_N$		174				(Sekino et al. 2006)
	D	Afa121 AB239647	$(AC)_nX_N(CA)_NT(AC)_NX_N(AC)_N(ACGC)_N(AC)_N$		244				(Sekino et al. 2006)
	D	Afa123 AB239648	$(AC)_N$		215				(Sekino et al. 2006)
	D	Afa125 AB239649	$(CA)_N(AC)_N$		198				(Sekino et al. 2006)
	D	Afa126 AB239650	$(CA)_N$		211				(Sekino et al. 2006)
	D	Afa129 AB239651	$(AC)_N(TCAC)_N(AC)_N$		208				(Sekino et al. 2006)
	D	Afa130 AB239652	$(CACG)_N(CA)_N$		219				(Sekino et al. 2006)
	D	Afa136 AB23953	$(GC)_N(AC)_N(GCAC)_NG(CA)_N$		213				(Sekino et al. 2006)
	D	Afa140B AB239654	$(AC)_n(ATACAC)_N$		214				(Sekino et al. 2006)
	D	Afa142 AB239655	$(CACG)_N(CATA)_N(CA)_NX_N(GCAC)_NA(CG)_N(CGCA)_NX_N(AC)_N$		230				(Sekino et al. 2006)
	D	Afa144 AB239656	$(CA)_NX_N(CG)_N(CA)_N$		175				(Sekino et al. 2006)
	D	Afa145 AB23957	$(AC)_N$		182				(Sekino et al. 2006)
	D	Afa147 AB239658	$(CA)_N(CGCA)_N$		233				(Sekino et al. 2006)
	D	Afa148	$(GCACACACGCAC)_NG(C$		206				(Sekino et al. 2006)

		AB239659	$GCA)_NX_N(CA)_N$						
	D	Afa149 AB239660	$(AC)_N$		226				(Sekino et al. 2006)
	D	Afa151 AB239661	$(AC)_N$		268				(Sekino et al. 2006)
	D	Afa153 AB239662	$(AC)_N$		188				(Sekino et al. 2006)
	D	Afa154 AB239663	$(AC)_N$		221				(Sekino et al. 2006)
	D	Afa155 AB239664	$(AC)_N(GCAC)_NGT(ACGC)_N(AC)_NG(CA)_N(CACG)_N(CA)_N(CACG)_N(CA)_n$		225				(Sekino et al. 2006)
	D	Afa158 AB239665	$(AC)_n$		219				(Sekino et al. 2006)
	D	Afa160 AB239666	$(ACGCAC)_NA(CACG)_N(CA)_N$		211				(Sekino et al. 2006)
	D	Afa162 AB239667	$(CA)_n$		235				(Sekino et al. 2006)
	D	Afa167 AB239668	$(AC)_NX_N(CA)_N$		231				(Sekino et al. 2006)
	D	Afa172 AB239669	$(AGACAC)_N(AGACAT)_N(AGACAC)_N(AC)_NX_N(AC)_n$		200				(Sekino et al. 2006)
	D	Afa174 AB239670	$(AC)_NG(CA)_nT(AC)_N$		199				(Sekino et al. 2006)
	D	Afa179 AB239671	$(CA)_N(CACG)_NX_N(GCACAC)_N(AC)_N$		204				(Sekino et al. 2006)
	D	Afa180 AB239672	$(CA)_n$		230				(Sekino et al. 2006)
	D	Afa182 AB239673	$(CA)_N(CGGA)_N$		216				(Sekino et al. 2006)
	D	Afa183 AB239674	$(AC)_NG(CGCA)_NX_N(AC)_N(AG)_N$		203				(Sekino et al. 2006)
	D	Afa185A AB239675	$(TG)_N$		198				(Sekino et al. 2006)
	D	Afa186 AB239676	$(CA)_N(CGCA)_N(CA)_NTA(CACG)_N$		202				(Sekino et al. 2006)
	D	Afa187 AB239709	$(AC)_N$		230				(Sekino et al. 2006)
	D	Afa188 AB239710	$(GCAC)_N(AC)_N(AT)_N$		177				(Sekino et al. 2006)

	D	Afa190 AB239711	$(CACG)_N(CA)_NX_N(AC)_N$		183				(Sekino et al. 2006)
	D	Afa193 AB239712	$(TCAC)_NX_N(CACT)_NT(ACTC)_NX_N(CACT)_N(CA)_NTC(CACT)_NT(ACTC)_NX_N(CACT)_N X_N(CA)_N TC(CACT)_N$		260				(Sekino et al. 2006)
	D	Afa194 AB239713	$(TC)_N$		179				(Sekino et al. 2006)
	D	Afa195 AB239714	$(CT)_N$		197				(Sekino et al. 2006)
	D	Afa203 AB239715	$(TCAC)_N$		184				(Sekino et al. 2006)
	D	Afa207 AB239716	$(CTCA)_N$		182				(Sekino et al. 2006)
	D	Afa208 AB239717	$(CT)_N(CA)_N$		196				(Sekino et al. 2006)
	D	Afa209 AB239718	$(ACTC)_NA(CACT)_N$		176				(Sekino et al. 2006)
	D	Afa212 AB239719	$(TG)_NX_N(TG)_N(TC)_NX_N(TG)_N(AG)_N$		192				(Sekino et al. 2006)
Indo-Pacific									
H. asinina									
	D	Hau2 G62416	$(CA)_n$	40	166–168	2	–	0.29	(Selvamani et al. 2000)
	D	Hau9 G62417	$(CA)_n$	40	124–134	6	–	0.65	(Selvamani et al. 2000)
	D	Hau10 G62418	$(CA)_n(GA)_n$	41	140–170	14	–	0.87	(Selvamani et al. 2000)
	D	Hau13 G62419	$(CA)_n$	41	128–182	25	–	0.96	(Selvamani et al. 2000)
	D	Hau1M G62220	$(CA)_n$	37	94–136	17	–	0.92	(Selvamani et al. 2000)
	D	Hau2J G62216	$(CA)_N(CT)_N$	35	235–265	14	–	0.91*	(Selvamani et al. 2000)
	D	Hau2K G62221	$(CA)_N$	38	102–158	22	–	0.92	(Selvamani et al. 2000)
	D	Hau2L G62217	$(CA)_N(AG)_N$	38	198–232	16	–	0.93*	(Selvamani et al. 2000)
	D	Hau3C G62219	$(CA)_N$	22	122–140	9	–	0.82*	(Selvamani et al. 2000)

	D	Hau3D G62222	$(CA)_N(CG)_N(CA)_N$	21	192–238	16	–	0.93	(Selvamani et al. 2000)
	D	Hau3E G62218	$(CA)_NTA(CA)_NTA(CA)_N$ TACATA $(CA)_NTA(CA)_NTA(CA)_N$ TACACATA $(CA)_NTACATA(CA)_NTA(CA)_N$ $TA(CA)_N$ $(TA)_N$	40	186–230	12	–	0.77	(Selvamani et al. 2000)
		CUHas1	$(GT)_NN_N(GT)_N$	72	258–360	26	0.85	0.93	(Tang et al. 2004)
		CUHas2	$(AT)_N(GT)_N$	65	286–340	21	0.68	0.93	(Tang et al. 2004)
		CUHas3	$(GT)_N(GA)_N$	71	134–178	13	0.62	0.82	(Tang et al. 2004)
		CUHas4	$(GT)_N(TGCA)_NN_N(GT)_N$	67	222–250	5	0.4	0.59	(Tang et al. 2004)
		CUHas5	$(GT)_N$	72	104–173	19	0.35	0.91	(Tang et al. 2004)
		CUHas6	$(GT)_N$	48	232–240	6	0.75	0.71	(Tang et al. 2004)
		CUHas7	$(ACGC)_N$	48	112–126	3	0.27	0.24	(Tang et al. 2004)
		CUHas8	$(AGTG)_N$	72	148–238	19	0.71	0.88	(Tang et al. 2004)
		CUHas9	$(GT)_N$	48	148–240	26	0.81	0.92	(Tang et al. 2004)
		CUHas10	$(CA)_n$	48	118–160	9	0.42	0.63	(Tang et al. 2004)
		CUHas1		81		26	–	–	(Tang et al. 2005)
		CUHas4		81		5	–	–	(Tang et al. 2005)
		CUHas5		81		23	–	–	(Tang et al. 2005)
Southern Africa									
H. midae	A	CmrHr 2.15 AF195956		199		19	0.429	0.669*	(Evans et al. 2004b)
	A	CmrHr 2.23 AF302832		205		2	0.188	0.243	(Evans et al. 2004b)
	A	CmrHr 2.29 AF302834		205		17	0.554	0.674*	(Evans et al. 2004b)
	D	HmD14 AY303333	$(CA)_N$	27	142–180	16	0.67	0.76	(Bester et al. 2004)
	D	HmD33 AY303334	$(GAGT)_n$	22	129–205	11	0.32	0.87*	(Bester et al. 2004)
	D	HmD36 AY303335	$(GTGA)_N$	23	220–304	15	0.43	0.89*	(Bester et al. 2004)
	D	HmD55 AY303337	$(GTGA)_N$	32	183–211	9	0.68	0.80	(Bester et al. 2004)
	D	HmD59 AY303338	$(CA)_N$	32	106–150	15	0.78	0.84	(Bester et al. 2004)
	D	HmD60 AY303339	$(CA)_N$	14	155–171	8	0.14	0.86*	(Bester et al. 2004)

	D	HmD61 AY303340	$(CA)_N$	28	234–298	11	0.61	0.82	(Bester et al. 2004)
	D	HmD11 AY303341	$(TCTG)_N$	30	292–352	5	0.32	0.66	(Bester et al. 2004)
	D	HmD30 AY303342	$(AGTC)_n$	27	124–150	11	0.7	0.80	(Bester et al. 2004)
	D	HmSP1 AY303346	$(CA)_N(CGCA)_N(CA)_N$	21	192–276	21	0.48	0.93*	(Bester et al. 2004)
	D	HmSP5 AY303344	$(AC)_N$	31	185–219	14	0.63	0.74	(Bester et al. 2004)
Australia									
H. rubra									
	D	CmrHr1.11 AF194951	$(AC)_N$	14	172–176	2	0.43	0.41^{NA}	(Evans et al. 2000)
	D	CmrHr1.14 AF194952	$(GT)_n$	31	251–275	4	0.25	0.40^{NA}	(Evans et al. 2000)
	D	CmrHr1.24 AF194953	$(AT)_N$	27	222–228	4	0.28	0.50^{NA}	(Evans et al. 2000)
	D	CmrHr1.25 AF194954	$(CA)_N(AT)_n(TG)_N$	14	291–309	9	0.14	0.83^{NA}	(Evans et al. 2000)
	D	CmrHr2.9 AF194956	$(GT)_N$	14	159–233	13	0.43	0.87^{NA}	(Evans et al. 2000)
	D	CmrHr2.14 AF194957	$(GAGT)_n$	17	199–237	8	0.76	0.79^{NA}	(Evans et al. 2000)
	D	CmrHr2.26 AF194958	$(ATTC)_NT_NC(ATTC)_N$	15	190–212	8	0.6	0.83^{NA}	(Evans et al. 2000)
	D	CmrHr2.30 AF194959	$(GT)_n(TG)_N(AG)_N(TG)_n$	16	284–328	16	0.6	0.90^{NA}	(Evans et al. 2000)
	D	CmrHr2.36 AF194960	$(AC)_N$	15	83–121	8	0.46	0.74^{NA}	(Evans et al. 2000)
	D, C	CmrHr1.5 AF302824	$(CAGA)_N$		126				(Evans et al. 2001)
	D, C	CmrHr1.6 AF302825	$(CA)_n$		89				(Evans et al. 2001)
	D, C	CmrHr1.23 AF302826	$(AC)_N$		122				(Evans et al. 2001)
	D, C	CmrHr2.3 AF302827	$(GT)_NTT(TG)_N$		100				(Evans et al. 2001)
	D, C	CmrHr2.5 AF194955	$(GT)_N$		283–299				(Evans et al. 2001)

	D, C	CmrHr2.15 AF194956	$(CA)_N$		288			(Evans et al. 2001)
	D, C	CmrHr2.17 AF302828	$(GT)_N$		226			(Evans et al. 2001)
	D, C	CmrHr2.18 AF302829	$(GAGT)_N$		134			(Evans et al. 2001)
	D, C	CmrHr2.20 AF302830	$(AC)_N(GCAC)_N$		186			(Evans et al. 2001)
	D, C	CmrHr2.22 AF302831	$(CA)_N$		117–193			(Evans et al. 2001)
	D, C	CmrHr2.23 AF302832	$(AC)_N$		258–266			(Evans et al. 2001)
	D, C	CmrHr2.27 AF302833	$(GT)_N(GCGT)_N(GT)_N$		347			(Evans et al. 2001)
	D, C	CmrHr2.29 AF302834	$(CA)_N$		321			(Evans et al. 2001)
	C	CmrHr1.11						(Evans et al. 2001)
	C	CmrHr1.14						(Evans et al. 2001)
	C	CmrHr1.24						(Evans et al. 2001)
	C	CmrHr1.25						(Evans et al. 2001)
	C	CmrHr2.9						(Evans et al. 2001)
	C	CmrHr2.14						(Evans et al. 2001)
	C	CmrHr2.26						(Evans et al. 2001)
	C	CmrHr2.30						(Evans et al. 2001)
	C	CmrHr2.36						(Evans et al. 2001)
	D	RubGT1 AF027572	$(GT)_N$	100		41		* (Huang and Hanna 1998)[A]
	D	RubCA1 AF027573	$(CA/G)_N(CA)_N$	100		30		* (Huang and Hanna 1998)[A]
	D	RubGACA1 AF027574	$(GACA)_N$	100		8		* (Huang and Hanna 1998)[A]
	C	RubGT1 AF027572		100		41	0.37	0.955^{NA} (Huang et al. 2000)[A]
	C	RubCA1 AF027573		100		30	0.38	0.955^{NA} (Huang et al. 2000)[A]
	C	RubGACA1 AF027574		100		8	0.19	0.814^{NA} (Huang et al. 2000)[A]
	D	CmrHr1.14		35	251–289	–	–	– (Conod et al. 2002)
	D	CmrHr2.14		36	199–241	–	–	– (Conod et al. 2002)
	D	CmrHr2.26		60	168–296	–	–	– (Conod et al. 2002)

	D	CmrHr2.30		194	282–402	–	–	–	(Conod et al. 2002)
	D	RubCA1		151	110–208	–	–	–	(Conod et al. 2002)
	D	Hrub1.D03 DQ277991	$(AC)_N$	10	160–205	8	0.900	–	(Baranski et al. 2006)
	D	Hrub1.D04 DQ277992	$(GT)_N$	11	238–328	17	0.818	–	(Baranski et al. 2006)
	D	Hrub1.D12 DQ277993	$(GT)_N(GTCT)_N$	11	194–212	12	0.818	–	(Baranski et al. 2006)
	D	Hrub1.G04 DQ277994	$(GT)_N(GCGT)_N$	9	190–220	10	0.556	–	(Baranski et al. 2006)
	D	Hrub1.H05 DQ277995	$(AGAC)_N$	10	150–180	6	0.400	–	(Baranski et al. 2006)
	D	Hrub1.H07 DQ277996	$(ACTC)_N$	10	149–157	3	0.300	–	(Baranski et al. 2006)
	D	Hrub1.H08 DQ277997	$(AC)_N$	9	152–219	9	0.667	–	(Baranski et al. 2006)
	D	Hrub10.B09 DQ277998	$(GT)_n$	9	239–323	7	0.333	–	(Baranski et al. 2006)
	D	Hrub10.B10 DQ277999	$(GT)_n$	9	252–308	7	0.222	–	(Baranski et al. 2006)
	D	Hrub10.B11 DQ278000	$(AC)_N$	9	198–221	9	0.778	–	(Baranski et al. 2006)
	D	Hrub10.E02 DQ278001	$(TG)_n$	9	143–215	11	0.889	–	(Baranski et al. 2006)
	D	Hrub10.G02 DQ278002	$(TG)_n$	9	242–309	15	1.000	–	(Baranski et al. 2006)
	D	Hrub10.G10 DQ278003	$(TG)_n$	9	203–220	9	0.889	–	(Baranski et al. 2006)
	D	Hrub10.G12 DQ278004	$(GT)_n(GA)_N$	9	152–170	8	1.000	–	(Baranski et al. 2006)
	D	Hrub10.H09 DQ278005	$(CA)_n$	9	149–253	14	1.000	–	(Baranski et al. 2006)
	D	Hrub10.H10 DQ278006	$(TG)_N(TC)_N(TG)_n(TC)_N$	9	201–271	14	0.889	–	(Baranski et al. 2006)
	D	Hrub11.A02 DQ278007	$(TG)_N(CG)_N$	8	154–304	9	0.250	–	(Baranski et al. 2006)
	D	Hrub11.A05 DQ278008	$(GC)_n(GTGC)_N(GT)_N$	8	210–259	7	0.875	–	(Baranski et al. 2006)
	D	Hrub11.A07 DQ278009	$(TG)_n$	9	77–101	8	0.889	–	(Baranski et al. 2006)

	D	Hrub11.A10 DQ278010	$(GT)_n$	9	218–261	8	0.444	–	(Baranski et al. 2006)
	D	Hrub11.A12 DQ278011	$(GT)_N(G)_N$	4	205–209	4	0.000	–	(Baranski et al. 2006)
	D	Hrub11.B04 DQ278012	$(TGGA)_n$	8	222–303	8	0.500	–	(Baranski et al. 2006)
	D	Hrub11.B09 DQ278013	$(GT)_N$	9	218–270	6	0.778	–	(Baranski et al. 2006)
	D	Hrub11.D03 DQ278014	$(AC)_n$	11	229–275	9	0.727	–	(Baranski et al. 2006)
	D	Hrub11.D08 DQ278015	$(GT)_n$	9	176–198	9	0.889	–	(Baranski et al. 2006)
	D	Hrub11.E05 DQ278016	$(GTT)_N$	6	201–210	4	0.500	–	(Baranski et al. 2006)
	D	Hrub12.A02 DQ278017	$(CA)_N$	9	202–219	4	0.333	–	(Baranski et al. 2006)
	D	Hrub12.A09 DQ278018	$(GT)_n$	8	193–306	11	0.750	–	(Baranski et al. 2006)
	D	Hrub12.A11 DQ278019	$(TTG)_n$	9	236–295	9	0.556	–	(Baranski et al. 2006)
	D	Hrub12.B10 DQ278020	$(CAA)_n$	9	205–250	8	1.000	–	(Baranski et al. 2006)
	D	Hrub12.D02 DQ278021	$(AC)_N$	9	209–310	5	0.333	–	(Baranski et al. 2006)
	D	Hrub12.E01 DQ278022	$(AGA)_N$	9	175–190	5	0.111	–	(Baranski et al. 2006)
	D	Hrub12.E07 DQ278023	$(GA)_N(GT)_N$	11	140–210	17	0.818	–	(Baranski et al. 2006)
	D	Hrub12.E10 DQ278024	$(GATG)_N$	11	280–314	8	0.909	–	(Baranski et al. 2006)
	D	Hrub12.E12 DQ278025	$(CA)_n(CT)_N$	7	185–255	7	0.571	–	(Baranski et al. 2006)
	D	Hrub12.F05 DQ278026	$(GT)_n$	9	193–231	9	0.667	–	(Baranski et al. 2006)
	D	Hrub12.F06 DQ278027	$(GA)_N(GTTT)_n(GT)_N$	9	160–293	11	0.556	–	(Baranski et al. 2006)
	D	Hrub12.F09 DQ278028	$(TG)_n$	7	266–319	7	0.714	–	(Baranski et al. 2006)
	D	Hrub12.G01 DQ278029	$(GT)_n$	9	243–307	11	1.000	–	(Baranski et al. 2006)

	D	Hrub12.G07 DQ278030	$(TC)_N(AC)_n$	9	194–299	13	0.778	–	(Baranski et al. 2006)
	D	Hrub12.H06 DQ278031	$(CA)_n$	6	178–195	4	0.000	–	(Baranski et al. 2006)
	D	Hrub13.A02 DQ278032	$(AC)_N$	4	235–250	6	0.500	–	(Baranski et al. 2006)
	D	Hrub13.B04 DQ278033	$(TG)_N$	3	282–320	5	1.000	–	(Baranski et al. 2006)
	D	Hrub13.C11 DQ278034	$(TG)_N$	3	251–283	4	0.667	–	(Baranski et al. 2006)
	D	Hrub13.C12 DQ278035	$(TGC)_N(GTT)_N$	7	220–223	2	0.143	–	(Baranski et al. 2006)
	D	Hrub13.E07 DQ278036	$(ACA)_N$	6	210–225	2	0.333	–	(Baranski et al. 2006)
	D	Hrub13.F06 DQ278037	$(GT)_N$	6	217–224	4	0.500	–	(Baranski et al. 2006)
	D	Hrub13.F07 DQ278038	$(GT)_n$	4	265–345	3	0.500	–	(Baranski et al. 2006)
	D	Hrub13.G01 DQ278039	$(TG)_n$	6	231–262	6	0.667	–	(Baranski et al. 2006)
	D	Hrub13.H05 DQ278040	$(ACAG)_n$	6	233–309	8	1.167[B]	–	(Baranski et al. 2006)
	D	Hrub13.H11 DQ278041	$(CACG)_n(CA)_N$	7	235–308	7	0.714	–	(Baranski et al. 2006)
	D	Hrub14.A02 DQ278042	$(GT)_n$	6	233–269	8	0.833	–	(Baranski et al. 2006)
	D	Hrub14.A04 DQ278043	$(TG)_N$	7	225–253	0	0.857	–	(Baranski et al. 2006)
	D	Hrub14.A10 DQ278044	$(ATGT)_N(GT)_N$	6	266–350	0	1.000	–	(Baranski et al. 2006)
	D	Hrub15.A01 DQ278045	$(CAGA)_n$	7	255–269	3	0.429	–	(Baranski et al. 2006)
	D	Hrub16.C02 DQ278046	$(CA)_n$	6	262–290	8	0.833	–	(Baranski et al. 2006)
	D	Hrub16.D06 DQ278047	$(TG)_N$	6	167–223	6	0.500	–	(Baranski et al. 2006)
	D	Hrub16.F04 DQ278048	$(GT)_n(GC)_N(GT)_N$	7	208–248	8	0.857	–	(Baranski et al. 2006)
	D	Hrub16.F06 DQ278049	$(GT)_N(GCGT)_N(GT)_n$	6	192–266	7	0.833	–	(Baranski et al. 2006)

	D	Hrub16.F08 DQ278050	$(AT)_n$	6	280–282	3	0.167	–	(Baranski et al. 2006)
	D	Hrub16.G01 DQ278051	$(TG)_n$	6	256–330	1	0.833	–	(Baranski et al. 2006)
	D	Hrub16.G08 DQ278052	$(GT)_N(GCGT)_N(GT)_n$	6	223–275	1	1.000	–	(Baranski et al. 2006)
	D	Hrub17.D11 DQ278053	$(TCCA)_n$	7	240–252	2	0.143	–	(Baranski et al. 2006)
	D	Hrub17.E04 DQ278054	$(TG)_n$	6	214–234	6	0.667	–	(Baranski et al. 2006)
	D	Hrub17.E12 DQ278055	$(GT)_n(GCGT)_n$	6	248–263	4	0.833	–	(Baranski et al. 2006)
	D	Hrub17.F05 DQ278056	$(TG)_n$	6	267–327	5	0.833	–	(Baranski et al. 2006)
	D	Hrub2.B01 DQ278057	$(AAC)_N$	11	151–172	9	0.909	–	(Baranski et al. 2006)
	D	Hrub2.B05 DQ278058	$(GC)_N(AC)_N$	9	223–247	8	0.667	–	(Baranski et al. 2006)
	D	Hrub2.D04 DQ278059	$(GT)_N(GTGC)_N(GC)_n$	11	269–303	0	0.545	–	(Baranski et al. 2006)
	D	Hrub2.G01 DQ278060	$(GT)_n$	10	150–164	6	0.800	–	(Baranski et al. 2006)
	D	Hrub2.H01 DQ278061	$(ACAG)_n$	11	262–346	8	0.545	–	(Baranski et al. 2006)
	D	Hrub3.A08 DQ278062	$(CA)_N$	10	269–285	8	0.500	–	(Baranski et al. 2006)
	D	Hrub3.B04 DQ278063	$(GCGT)_n$	11	240–256	5	0.545	–	(Baranski et al. 2006)
	D	Hrub3.E02 DQ278064	$(GT)_N$	11	244–274	10	0.909	–	(Baranski et al. 2006)
	D	Hrub3.F01 DQ278065	$(T)_N(GT)_n$	10	205–264	7	0.600	–	(Baranski et al. 2006)
	D	Hrub3.F03 DQ278066	$(CAG)_N(CAA)_n$	8	281–306	6	0.375	–	(Baranski et al. 2006)
	D	Hrub3.F11 DQ278067	$(CA)_n(CGCA)_n(CA)_N(CT)_n$	8	224–251	9	0.875	–	(Baranski et al. 2006)
	D	Hrub3.G06 DQ278068	$(AC)_n$	9	267–344	11	0.667	–	(Baranski et al. 2006)
	D	Hrub4.A02 DQ278069	$(AC)_N$	9	239–289	13	0.556	–	(Baranski et al. 2006)

	D	Hrub4.A03 DQ278070	$(GT)_N(GC)_N$	11	243–315	9	0.455	–	(Baranski et al. 2006)
	D	Hrub4.A10 DQ278071	$(GT)_n$	8	220–305	10	0.750	–	(Baranski et al. 2006)
	D	Hrub4.B09 DQ278072	$(TA)_N(TG)_n(TGTA)_N$	8	281–360	9	0.125	–	(Baranski et al. 2006)
	D	Hrub4.E05 DQ278073	$(CA)_N$	8	206–216	5	0.875	–	(Baranski et al. 2006)
	D	Hrub4.E06 DQ278074	$(CA)_N$	9	207–239	11	1.000	–	(Baranski et al. 2006)
	D	Hrub4.F07 DQ278075	$(ATGG)_N$	11	231–248	5	0.636	–	(Baranski et al. 2006)
	D	Hrub4.G05 DQ278076	$(TG)_n(G)_N$	9	183–193	6	0.111	–	(Baranski et al. 2006)
	D	Hrub4.H03 DQ278077	$(GT)_N$	11	184–255	8	0.909	–	(Baranski et al. 2006)
	D	Hrub4.H09 DQ278078	$(GT)_N$	11	239–276	13	0.909	–	(Baranski et al. 2006)
	D	Hrub4.H11 DQ278079	$(AAG)_n$	11	241–267	9	0.818	–	(Baranski et al. 2006)
	D	Hrub4.H12 DQ278080	$(CA)_N$	8	261–293	7	0.500	–	(Baranski et al. 2006)
	D	Hrub6.A05 DQ278081	$(GATG)_n$	9	225–267	7	0.444	–	(Baranski et al. 2006)
	D	Hrub6.A10 DQ278082	$(CACG)_N(CA)_N$	11	214–296	8	0.455	–	(Baranski et al. 2006)
	D	Hrub6.C04 DQ278083	$(CTGT)_N$	10	215–239	5	0.700	–	(Baranski et al. 2006)
	D	Hrub6.E06 DQ278084	$(TG)_n$	9	183–266	11	0.444	–	(Baranski et al. 2006)
	D	Hrub6.G09 DQ278085	$(CAGA)_n$	11	76–80	2	0.000	–	(Baranski et al. 2006)
	D	Hrub7.A05 DQ278086	$(TGT)_n$	9	230–273	11	0.889	–	(Baranski et al. 2006)
	D	Hrub7.B11 DQ278087	$(AAC)_N(AC)_N(TC)_n$	11	237–300	11	0.909	–	(Baranski et al. 2006)
	D	Hrub7.C06 DQ278088	$(GAGT)_N(GT)_N(GA)_N(GAGT)_n$	11	167–228	9	0.545	–	(Baranski et al. 2006)
	D	Hrub7.D01 DQ278089	$(TG)_n$	10	235–274	8	0.600	–	(Baranski et al. 2006)

	D	Hrub7.F02 DQ278090	$(TG)_N$	10	250–294	9	0.500	–	(Baranski et al. 2006)
	D	Hrub7.G05 DQ278091	$(GA)_n(GT)_n$	9	112–266	12	0.667	–	(Baranski et al. 2006)
	D	Hrub7.G10 DQ278092	$(AC)_N$	11	204–234	8	0.545	–	(Baranski et al. 2006)
	D	Hrub7.H10 DQ278093	$(CA)_n$	11	175–217	13	0.364	–	(Baranski et al. 2006)
	D	Hrub8.A03 DQ278094	$(CA)_N$	6	142–183	8	0.833	–	(Baranski et al. 2006)
	D	Hrub8.A09 DQ278095	$(GA)_N(GT)_N(AG)_n$	9	174–235	11	0.778	–	(Baranski et al. 2006)
	D	Hrub8.D01 DQ278096	$(TC)_n(GT)_n(GCGT)_n$	9	278–326	10	0.444	–	(Baranski et al. 2006)
	D	Hrub8.D02 DQ278097	$(ATGG)_n$	6	152–204	8	0.667	–	(Baranski et al. 2006)
	D	Hrub8.F05 DQ278098	$(CA)_n$	11	207–277	15	0.909	–	(Baranski et al. 2006)
	D	Hrub8.F11 DQ278099	$(CA)_n$	6	105–122	3	0.333	–	(Baranski et al. 2006)
	D	Hrub8.G12 DQ278100	$(AC)_N$	11	238–261	6	0.727	–	(Baranski et al. 2006)
	D	Hrub9.A04 DQ278101	$(ATGG)_n$	9	236–256	5	0.556	–	(Baranski et al. 2006)
	D	Hrub9.A09 DQ278102	$(GT)N(GA)_N$	9	182–242	11	1.000	–	(Baranski et al. 2006)
	D	Hrub9.B04 DQ278103	$(GT)N(GA)_N$	9	247–275	10	0.889	–	(Baranski et al. 2006)
	D	Hrub9.B05 DQ278104	$(TG)_N$	11	184–215	10	0.727	–	(Baranski et al. 2006)
	D	Hrub9.C01 DQ278105	$(AC)_N$	6	284–292	5	0.500	–	(Baranski et al. 2006)
	D	Hrub9.C09 DQ278106	$(GACA)_N(AC)_n$	9	167–197	10	0.222	–	(Baranski et al. 2006)
	D	Hrub9.C11 DQ278107	$(GT)_N$	7	236–258	8	0.857	–	(Baranski et al. 2006)
	D	Hrub9.E04 DQ278108	$(AC)_n$	11	203–230	8	0.909	–	(Baranski et al. 2006)
	D	Hrub9.F09 DQ278109	$(CA)_N$	8	224–257	9	0.250	–	(Baranski et al. 2006)

	D	Hrub9.F11 DQ278110	$(GT)_n$	9	205–339	16	0.889	–	(Baranski et al. 2006)
	D	Hrub9.G01 DQ278111	$(GT)_N$	11	209–246	9	0.818	–	(Baranski et al. 2006)
	D	Hrub9.H03 DQ278112	$(CACT)_n$	9	200–228	3	0.444	–	(Baranski et al. 2006)
	D	Hrub9.H06 DQ278113	$(CA)_n$	9	131–186	8	0.111	–	(Baranski et al. 2006)
	D	Hrub9.H08 DQ278114	$(AC)_N$	9	283–307	11	1.000	–	(Baranski et al. 2006)
	D	Hrub9.H11 DQ278115	$(GACA)_N$	9	256–281	7	0.778	–	(Baranski et al. 2006)
	A	CmrHr1.14		540		–	–	–	(Temby et al. 2007)
	A	CmrHr1.24		540		–	–	–	(Temby et al. 2007)
	A	CmrHr2.14		540		–	–	–	(Temby et al. 2007)

Table A5.2: Microsatellite cross amplification in *Haliotis* spp. Listed are the original species for which a microsatllite locus was developed, the name of the locus as labeled in Table A5.1, and the reference for its isolation, the species for which the microsatellite primers successfully amplified a product, species for which the microsatellite primers failed to amplify a product, species for which further optimization of the microsatellite were tested (NA, further optimization was not reported; Successful, further optimization was attempted and produced a usable microsatellite locus in the new species; Unsuccessful, further optimization was attempted and failed to produce a usable microsatellite locus), and the reference for the study testing cross amplification. Colored text refers to locations in Figure 1.1. Black text was used instead of yellow for species located in the North Pacific ocean.

Original Species	**Locus**	**Successful cross-amplification**	**Unsuccessful cross-amplification**	**Further optimization**	**Reference**
H. fulgens	Hful240 (Cruz et al. 2005)		*H. rufescens* *H. corrugata*	NA	(Cruz et al. 2005) (Cruz et al. 2005)
	Hful333 (Cruz et al. 2005)		*H. rufescens* *H. corrugata*	NA	(Cruz et al. 2005) (Cruz et al. 2005)
	Hful369 (Cruz et al. 2005)	*H. rufescens* *H. corrugata*		NA	(Cruz et al. 2005) (Cruz et al. 2005)
	Hful442 (Cruz et al. 2005)	*H. rufescens* *H. corrugata*		NA	(Cruz et al. 2005) (Cruz et al. 2005)
	Hful547 (Cruz et al. 2005)		*H. rufescens* *H. corrugata*	NA	(Cruz et al. 2005) (Cruz et al. 2005)
	Hful603 (Cruz et al. 2005)	*H. rufescens* *H. corrugata*		NA	(Cruz et al. 2005) (Cruz et al. 2005)
	Hful910 (Cruz et al. 2005)	*H. rufescens* *H. corrugata*		NA	(Cruz et al. 2005) (Cruz et al. 2005)
	Hful1136 (Cruz et al. 2005)	*H. rufescens* *H. corrugata*		NA	(Cruz et al. 2005) (Cruz et al. 2005)
H. kamtschatkana	Hka3 (Miller et al. 2001)	*H. sorenseni* *H. rufescens* *H. corrugata*		Successful Successful	(Gruenthal and Burton 2005) (Gruenthal et al. 2007) (Díaz-Viloria et al. 2008)
	Hka28 (Miller et al. 2001)	*H. sorenseni* *H. rufescens* *H. fulgens*		Successful Successful Successful	(Gruenthal and Burton 2005) (Gruenthal et al. 2007) (Gutiérrez-Gonzalez and Perez-Enriquez 2005); (Gutiérrez-Gonzalez et al. 2007)
	Hka40 (Miller et al. 2001)	*H. sorenseni* *H. rufescens*		Successful Successful	(Gruenthal and Burton 2005) (Gruenthal et al. 2007)

	Hka56 (Miller et al. 2001)	*H. sorenseni* *H. rufescens* *H. fulgens* *H. corrugata*		Successful Successful Successful	(Gruenthal and Burton 2005) (Gruenthal et al. 2007) (Gutiérrez-Gonzalez and Perez-Enriquez 2005); (Gutiérrez-Gonzalez et al. 2007) (Díaz-Viloria et al. 2008)
	Hka80 (Miller et al. 2001)	*H. sorenseni* *H. rufescens*		Successful Successful	(Gruenthal and Burton 2005) (Gruenthal et al. 2007)
H. corrugata	Hco6 (Díaz-Viloria et al. 2008)	*H. fulgens* *H. rufescens*		Successful Successful	(Díaz-Viloria et al. 2008) (Díaz-Viloria et al. 2008)
	Hco15 (Díaz-Viloria et al. 2008)	*H. fulgens* *H. rufescens*		Successful Successful	(Díaz-Viloria et al. 2008) (Díaz-Viloria et al. 2008)
	Hco16 (Díaz-Viloria et al. 2008)		*H. fulgens* *H. rufescens*		(Díaz-Viloria et al. 2008) (Díaz-Viloria et al. 2008)
	Hco19 (Díaz-Viloria et al. 2008)	*H. fulgens* *H. rufescens*		Successful Successful	(Díaz-Viloria et al. 2008) (Díaz-Viloria et al. 2008)
	Hco22 (Díaz-Viloria et al. 2008)	*H. fulgens* *H. rufescens*		Successful Successful	(Díaz-Viloria et al. 2008) (Díaz-Viloria et al. 2008)
	Hco23 (Díaz-Viloria et al. 2008)		*H. fulgens* *H. rufescens*		(Díaz-Viloria et al. 2008) (Díaz-Viloria et al. 2008)
	Hc43 (Díaz-Viloria et al. 2008)	*H. fulgens* *H. rufescens*		Successful Successful	(Díaz-Viloria et al. 2008) (Díaz-Viloria et al. 2008)
	Hco47 (Díaz-Viloria et al. 2008)	*H. fulgens* *H. rufescens*		Successful Successful	(Díaz-Viloria et al. 2008) (Díaz-Viloria et al. 2008)
	Hco97 (Díaz-Viloria et al. 2008)	*H. fulgens* *H. rufescens*		Successful Successful	(Díaz-Viloria et al. 2008) (Díaz-Viloria et al. 2008)
	Hco194 (Díaz-Viloria et al. 2008)	*H. fulgens* *H. rufescens*		Successful Successful	(Díaz-Viloria et al. 2008) (Díaz-Viloria et al. 2008)

H. discus hannai	Hd527 (Hara and Sekino 2005)	*H. discus discus*		Successful	(Hara and Sekino 2005)
	Hd535 (Hara and Sekino 2005)	*H. discus discus*		Successful	(Hara and Sekino 2005)
	Hd601 (Hara and Sekino 2005)	*H. discus discus*		Successful	(Hara and Sekino 2005)
	Hd604 (Hara and Sekino 2005)	*H. discus discus*		Successful	(Hara and Sekino 2005)
	Hd680 (Hara and Sekino 2005)	*H. discus discus*		Successful	(Hara and Sekino 2005)
	Hd715 (Hara and Sekino 2005)	*H. discus discus*		Successful	(Hara and Sekino 2005)
	Hd724 (Hara and Sekino 2005)	*H. discus discus*		Successful	(Hara and Sekino 2005)
	Hd731 (Hara and Sekino 2005)	*H. discus discus*		Successful	(Hara and Sekino 2005)
H. asinina	CUHas1 (Tang et al. 2004)		*H. ovina* *H. varia*		(Tang et al. 2004)
H. asinina	CUHas2 (Tang et al. 2004)		*H. ovina* *H. varia*		(Tang et al. 2004)
H. asinina	CUHas3 (Tang et al. 2004)		*H. ovina* *H. varia*		(Tang et al. 2004)
H. asinina	CUHas4 (Tang et al. 2004)		*H. ovina* *H. varia*		(Tang et al. 2004)
H. asinina	CUHas5 (Tang et al. 2004)		*H. ovina* *H. varia*		(Tang et al. 2004)
H. asinina	CUHas6 (Tang et al. 2004)		*H. ovina* *H. varia*		(Tang et al. 2004)

H. asinina	CUHas7 (Tang et al. 2004)		*H. ovina* *H. varia*		(Tang et al. 2004)
H. asinina	CUHas8 (Tang et al. 2004)		*H. ovina* *H. varia*		(Tang et al. 2004)
H. asinina	CUHas9 (Tang et al. 2004)		*H. ovina* *H. varia*		(Tang et al. 2004)
H. asinina	CUHas10 (Tang et al. 2004)		*H. ovina* *H. varia*		(Tang et al. 2004)
Haliotis rubra	CmrHr1.5 (Evans et al. 2001)	 *H. conicopora* *H. roei* *H. spadicea*	*H. laevigata* *H. scalaris* *H. asinina* *H. iris* *H. australis* *H. virginea* *H. midae* *H. fulgens* *H. corrugata*	 NA NA NA	(Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001)
Haliotis rubra	CmrHr1.6 (Evans et al. 2001)	*H. laevigata* *H. scalaris* *H. conicopora* *H. midae* *H. spadicea*	 *H. roei* *H. asinina* *H. iris* *H. australis* *H. virginea* *H. fulgens* *H. corrugata*	Unsuccessful NA NA NA NA	(Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001)
Haliotis rubra	CmrHr1.23 (Evans et al. 2001)	 *H. scalaris* *H. asinina*	*H. laevigata* *H. conicopora* *H. roei* 	 NA NA	(Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001)

			H. iris		(Evans et al. 2001)
			H. australis		(Evans et al. 2001)
			H. virginea		(Evans et al. 2001)
		H. midae		Unsuccessful*	(Evans et al. 2001)
		H. spadicea		NA	(Evans et al. 2001)
			H. fulgens		(Evans et al. 2001)
			H. corrugata		(Evans et al. 2001)
Haliotis rubra	CmrHr2.3 (Evans et al. 2001)	*H. laevigata*		Unsuccessful	(Evans et al. 2001)
		H. scalaris		NA	(Evans et al. 2001)
		H. conicopora		NA	(Evans et al. 2001)
		H. roei		NA	(Evans et al. 2001)
		H. asinina		NA	(Evans et al. 2001)
		H. iris		NA	(Evans et al. 2001)
		H. australis		NA	(Evans et al. 2001)
		H. virginea		NA	(Evans et al. 2001)
		H. midae		Successful	(Evans et al. 2001)
		H. spadicea		NA	(Evans et al. 2001)
		H. fulgens		Unsuccessful	(Evans et al. 2001)
		H. corrugata		NA	(Evans et al. 2001)
Haliotis rubra	CmrHr2.5 (Evans et al. 2001)	*H. laevigata*			(Evans et al. 2001)
		H. scalaris		NA	(Evans et al. 2001)
		H. conicopora		NA	(Evans et al. 2001)
		H. roei		NA	(Evans et al. 2001)
		H. asinina		NA	(Evans et al. 2001)
			H. iris		(Evans et al. 2001)
			H. australis		(Evans et al. 2001)
		H. virginea		NA	(Evans et al. 2001)
		H. midae		NA	(Evans et al. 2001)
		H. spadicea		NA	(Evans et al. 2001)
		H. fulgens		NA	(Evans et al. 2001)
		H. corrugata		NA	(Evans et al. 2001)
Haliotis rubra	CmrHr2.15* (Evans et al. 2001)		*H. laevigata*		(Evans et al. 2001)
			H. scalaris		(Evans et al. 2001)
		H. conicopora		NA	(Evans et al. 2001)
		H. roei		NA	(Evans et al. 2001)
			H. asinina		(Evans et al. 2001)
			H. iris		(Evans et al. 2001)

		H. virginea *H. midae* *H. spadicea*	*H. australis* *H. fulgens* *H. corrugata*	 NA Successful NA NA	(Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001); (Evans et al. 2004b) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001)
Haliotis rubra	CmrHr2.17 (Evans et al. 2001)	*H. laevigata* *H. scalaris*	 *H. conicopora* *H. roei* *H. asinina* *H. iris* *H. australis* *H. virginea* *H. midae* *H. spadicea* *H. fulgens* *H. corrugata*	NA NA	(Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001)
Haliotis rubra	CmrHr2.18 (Evans et al. 2001)	*H. laevigata* *H. scalaris*	 *H. conicopora* *H. roei* *H. asinina* *H. iris* *H. australis* *H. virginea* *H. midae* *H. spadicea* *H. fulgens* *H. corrugata*	NA NA	(Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001)
Haliotis rubra	CmrHr2.20 (Evans et al. 2001)	*H. laevigata* *H. scalaris* *H. conicopora* *H. roei* *H. asinina* *H. iris* *H. australis*		Unsuccessful NA NA NA NA NA NA	(Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001)

		H. virginea *H. midae* *H. spadicea* *H. fulgens* *H. corrugata*		NA Unsuccessful NA NA	(Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001)
Haliotis rubra	CmrHr2.22 (Evans et al. 2001)	*H. laevigata* *H. conicopora* *H. roei* *H. asinina* *H. iris* *H. fulgens*	*H. scalaris* *H. australis* *H. virginea* *H. midae* *H. spadicea* *H. corrugata*	Unsuccessful NA NA NA NA NA	(Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001)
Haliotis rubra	CmrHr2.23 (Evans et al. 2001)	*H. laevigata* *H. scalaris* *H. conicopora* *H. roei* *H. iris* *H. australis* *H. virginea* *H. midae* 	 *H. asinina* *H. spadicea* *H. fulgens* *H. corrugata*	Successful NA NA NA NA NA NA Successful 	(Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001); (Evans et al. 2004b) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001)
Haliotis rubra	CmrHr2.27 (Evans et al. 2001)	*H. laevigata* *H. scalaris* *H. conicopora* *H. roei* *H. asinina* *H. iris* *H. australis* *H. virginea*		Unsuccessful NA NA NA NA NA NA NA	(Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001)

		H. midae *H. spadicea* *H. fulgens* *H. corrugata*		Unsuccessful NA Unsuccessful NA	(Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001)
Haliotis rubra	CmrHr2.29 (Evans et al. 2001)	*H. laevigata* *H. scalaris* *H. conicoopora* *H. roei* *H. midae*	 *H. asinina* *H. iris* *H. australis* *H. virginea* *H. spadicea* *H. fulgens* *H. corrugata*	Unsuccessful NA NA NA Successful	(Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001); (Evans et al. 2004b) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001)
Haliotis rubra	CmrHr1.11 (Evans et al. 2000)	 *H. conicopora* *H. australis*	*H. laevigata* *H. scalaris* *H. roei* *H asinina* *H iris* *H. virginea* *H. midae* *H. spadicea* *H. fulgens* *H. corrugata*	 NA NA	(Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001)
Haliotis rubra	CmrHr1.14 (Evans et al. 2000)	*H. laevigata* *H. conicopora* *H. roei*	 *H. scalaris* *H asinina* *H iris* *H. australis* *H. virginea* *H. midae*	Unsuccessful NA NA	(Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001)

			H. spadicea *H. fulgens* *H. corrugata*		(Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001)
Haliotis rubra	CmrHr1.24 (Evans et al. 2000)	*H. laevigata* *H. scalaris* *H. conicopora* *H. roei* *H asinina* *H iris* *H. australis* *H. virginea* *H. midae* *H. spadicea*	 *H. fulgens* *H. corrugata*	 NA NA NA NA NA NA NA Unsuccessful NA NA NA	(Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001)
Haliotis rubra	CmrHr1.25 (Evans et al. 2000)	*H. laevigata*	 *H. scalaris* *H. conicopora* *H. roei* *H. asinina* *H. iris* *H. australis* *H. virginea* *H. midae* *H. spadicea* *H. fulgens* *H. corrugata*	Unsuccessful	(Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001)
Haliotis rubra	CmrHr2.9* (Evans et al. 2000)	 *H. conicopora* *H. roei* *H. virginea* *H. midae*	*H. laevigata* *H. scalaris* *H. asinina* *H. iris* *H. australis* *H. spadicea*	 NA NA NA Successful	(Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001)

		H. fulgens *H. corrugata*		NA NA	(Evans et al. 2001) (Evans et al. 2001)
Haliotis rubra	CmrHr2.14 (Evans et al. 2000)	*H. laevigata* *H. scalaris* *H. conicopora* *H. roei* *H. virginea*	 *H. asinina* *H. iris* *H. australis* *H. midae* *H. spadicea* *H. fulgens* *H. corrugata*	Successful NA NA NA NA	(Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001)
Haliotis rubra	CmrHr2.26 (Evans et al. 2000)	 *H. roei*	*H. laevigata* *H. scalaris* *H. conicopora* *H. asinina* *H. iris* *H. australis* *H. virginea* *H. midae* *H. spadicea* *H. fulgens* *H. corrugata*	 NA	(Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001)
Haliotis rubra	CmrHr2.30 (Evans et al. 2000)	*H. laevigata* *H. scalaris* *H. conicopora* *H. roei* *H. asinina* *H. iris* *H. australis* *H. virginea* *H. midae* *H. spadicea* *H. fulgens*	 	Successful NA NA NA NA NA NA NA Successful 	(Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001)

			H. corrugata		(Evans et al. 2001)
Haliotis rubra	CmrHr2.36 (Evans et al. 2000)		*H. laevigata* *H. scalaris* *H. conicopora* *H. roei* *H. asinina*		(Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001) (Evans et al. 2001)
		H. iris		NA	(Evans et al. 2001)
			H. australis		(Evans et al. 2001)
			H. virginea		(Evans et al. 2001)
		H. midae		Successful	(Evans et al. 2001)
		H. spadicea		NA	(Evans et al. 2001)
			H. fulgens		(Evans et al. 2001)
			H. corrugata		(Evans et al. 2001)
Haliotis rubra	RubGT1 (Huang and Hanna 1998)		*H. laevigata*		(Huang and Hanna 1998)
		H. roei		NA	(Huang and Hanna 1998)
		H. scalaris		NA	(Huang and Hanna 1998)
		H. conicopora		NA	(Huang and Hanna 1998)
			H. discus discus		(Huang and Hanna 1998)
			H. discus hannai		(Huang and Hanna 1998)
		H. gigantea		NA	(Huang and Hanna 1998)
		H. sieboldi		NA	(Huang and Hanna 1998)
			H. diversicolor		(Huang and Hanna 1998)
			H. pervum		(Huang and Hanna 1998)
			H. midae		(Huang and Hanna 1998)
			H. rufescens		(Huang and Hanna 1998)
			H. cracherodii		(Huang and Hanna 1998)
	RubCA1 (Huang and Hanna 1998)		*H. laevigata*		(Huang and Hanna 1998)
			H. roei		(Huang and Hanna 1998)
		H. scalaris		NA	(Huang and Hanna 1998)
		H. conicopora		NA	(Huang and Hanna 1998)
			H. discus discus		(Huang and Hanna 1998)
			H. discus hannai		(Huang and Hanna 1998)
			H. gigantea		(Huang and Hanna 1998)
			H. sieboldi		(Huang and Hanna 1998)
			H. diversicolor		(Huang and Hanna 1998)
			H. pervum		(Huang and Hanna 1998)
			H. midae		(Huang and Hanna 1998)

			H. rufescens *H. cracherodii*		(Huang and Hanna 1998) (Huang and Hanna 1998)
	RubGACA1 (Huang and Hanna 1998)	 *H. roei* *H. conicopora*	*H. laevigata* *H. scalaris* *H. discus discus* *H. discus hannai* *H. gigantea* *H. sieboldi* *H. diversicolor* *H. pervum* *H. midae* *H. rufescens* *H. cracherodii*	 NA NA	(Huang and Hanna 1998) (Huang and Hanna 1998) (Huang and Hanna 1998) (Huang and Hanna 1998) (Huang and Hanna 1998) (Huang and Hanna 1998) (Huang and Hanna 1998) (Huang and Hanna 1998) (Huang and Hanna 1998) (Huang and Hanna 1998) (Huang and Hanna 1998) (Huang and Hanna 1998)
Haliotis rubra	Hrub1.D03 (Baranski et al. 2006)	*H. laevigata* *H. coccoradiata*		NA NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub1.D04 (Baranski et al. 2006)	*H. laevigata*	 *H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub1.D12 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub1.G04 (Baranski et al. 2006)	*H. laevigata*	 *H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub1.H05 (Baranski et al. 2006)	*H. laevigata*	 *H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub1.H07 (Baranski et al. 2006)	*H. laevigata*	 *H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub1.H08 (Baranski et al. 2006)	*H. laevigata*	 *H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub10.B09 (Baranski et al. 2006)	*H. laevigata*	 *H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub10.B10 (Baranski et al. 2006)	*H. laevigata*	 *H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub10.B11 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub10.E02	*H. laevigata*		NA	(Baranski et al. 2006)

	(Baranski et al. 2006)	*H. coccoradiata*		NA	(Baranski et al. 2006)
Haliotis rubra	Hrub10.G02 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub10.G10 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub10.G12 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub10.H09 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub10.H10 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub11.A02 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub11.A05 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub11.A07 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub11.A10 (Baranski et al. 2006)	*H. laevigata* *H. coccoradiata*		NA NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub11.A12 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub11.B04 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub11.B09 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub11.D03 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub11.D08 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub11.E05 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub12.A02 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub12.A09 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub12.A11		*H. laevigata*		(Baranski et al. 2006)

	(Baranski et al. 2006)		*H. coccoradiata*		(Baranski et al. 2006)
Haliotis rubra	Hrub12.B10 (Baranski et al. 2006)	*H. laevigata* *H. coccoradiata*		NA NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub12.D02 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub12.E01 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub12.E07 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub12.E10 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub12.E12 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub12.F05 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub12.F06 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub12.F09 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub12.G01 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub12.G07 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub12.H06 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub13.A02 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub13.B04 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub13.C11 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub13.C12 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub13.E07 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub13.F06	*H. laevigata*		NA	(Baranski et al. 2006)

	(Baranski et al. 2006)	*H. coccoradiata*		NA	(Baranski et al. 2006)
Haliotis rubra	Hrub13.F07 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub13.G01 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub13.H05 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub13.H11 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub14.A02 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub14.A04 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub14.A10 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub15.A01 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub16.C02 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub16.D06 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub16.F04 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub16.F06 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub16.F08 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub16.G01 (Baranski et al. 2006)	*H. laevigata* *H. coccoradiata*		NA NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub16.G08 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub17.D11 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub17.E04 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub17.E12	*H. laevigata*		NA	(Baranski et al. 2006)

	(Baranski et al. 2006)		*H. coccoradiata*		(Baranski et al. 2006)
Haliotis rubra	Hrub17.F05 (Baranski et al. 2006)	*H. laevigata*	 *H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub2.B01 (Baranski et al. 2006)	*H. laevigata* *H. coccoradiata*		NA NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub2.B05 (Baranski et al. 2006)	*H. laevigata*	 *H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub2.D04 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub2.G01 (Baranski et al. 2006)	*H. laevigata*	 *H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub2.H01 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub3.A08 (Baranski et al. 2006)	*H. laevigata*	 *H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub3.B04 (Baranski et al. 2006)	*H. laevigata*	 *H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub3.E02 (Baranski et al. 2006)	*H. laevigata*	 *H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub3.F01 (Baranski et al. 2006)	*H. laevigata*	 *H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub3.F03 (Baranski et al. 2006)	*H. laevigata*	 *H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub3.F11 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub3.G06 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub4.A02 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub4.A03 (Baranski et al. 2006)	*H. laevigata*	 *H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub4.A10 (Baranski et al. 2006)	*H. laevigata*	 *H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub4.B09 (Baranski et al. 2006)	*H. laevigata*	 *H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub4.E05		*H. laevigata*		(Baranski et al. 2006)

	(Baranski et al. 2006)		*H. coccoradiata*		(Baranski et al. 2006)
Haliotis rubra	Hrub4.E06 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub4.F07 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub4.G05 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub4.H03 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub4.H09 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub4.H11 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub4.H12 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub6.A05 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub6.A10 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub6.C04 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub6.E06 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub6.G09 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub7.A05 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub7.B11 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub7.C06 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub7.D01 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub7.F02 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub7.G05	*H. laevigata*		NA	(Baranski et al. 2006)

	(Baranski et al. 2006)	*H. coccoradiata*		NA	(Baranski et al. 2006)
Haliotis rubra	Hrub7.G10 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub7.H10 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub8.A03 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub8.A09 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub8.D01 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub8.D02 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub8.F05 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub8.F11 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub8.G12 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub9.A04 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub9.A09 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub9.B04 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub9.B05 (Baranski et al. 2006)	*H. laevigata* *H. coccoradiata*		NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub9.C01 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub9.C09 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub9.C11 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub9.E04 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub9.F09		*H. laevigata*		(Baranski et al. 2006)

	(Baranski et al. 2006)		*H. coccoradiata*		(Baranski et al. 2006)
Haliotis rubra	Hrub9.F11 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub9.G01 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub9.H03 (Baranski et al. 2006)	*H. laevigata*	*H. coccoradiata*	NA	(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub9.H06 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub9.H08 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)
Haliotis rubra	Hrub9.H11 (Baranski et al. 2006)		*H. laevigata* *H. coccoradiata*		(Baranski et al. 2006) (Baranski et al. 2006)

APPENDIX 6: *HALIOTIS IRIS* MICROSATELLITE LOCI

AB1R
CTCTGACAACTGGACGCGTCCATAATGGCGAATATCCACATTTTGCAACATCG
AAATGTATATTAAACATAGACAGTTGTTGCTGCCTTTATAATGGACGTTTGAA
CTATTTCTCTGAAATTGAGCCAAACTGGTCTAAACGTTACTAGGGTTTTGTTT
CTCT***CTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT***
CTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT
CTCATAAACACTCACGCGCAAGCACACATTGTTCCAATGTAAAGTTTTATTAT
CACCAAATCGTACTTACAGGTTTGCCATTGTTAATGTGCTACCTCGGGAGAAC
AGCACCACCAGGGTAGTAAACCTCGGTCAGTACCGGCAGANATTCATATTGT
AGCAGGTTTCATGCAGGTTCTTTGCCAATATACATAG

AB3R
GTNNCTNGGACAGTCTNGTTCACNTTGTTTACCACAGAACCNGGAGTCCNGA
CACTGAAGACACTCTCCAGATGGCGTCACACTGACACAGCGAGGTAGACACA
CTCCGTCAACACAGGTCACGTTGTTACCGTCGCAGTTGCTGCATGGTTAGTTT
GAGGATGCTGTGTTATGAATTAGTTTAACAATAATAGCTATACACTTACATTG
TGAAATACATCTGTAT***CTATATCTATCTATCTATCTATCTATCTATCTATCTATC***
TATCTATCTATCTATCTGTGCGTTCGGGCGTGTGTGT

AB5R
ACCGAGCTCGAATTCGGACTAACATCAACAAGGACAAGGTCCGAGTGGTGCT
GGGTGACAAGCGTGTTGGCACTTGGCAGGCACAGGGGAACGGTGTGTTCATG
TCATCACCTGCGCGGAGGTCGACAGCCGACACAGTCACCTGCAACACCACCT
TCGTCCACAGCCTCATCGATGGTGGAATGTATCTCTCACACAAGAAGTTTAAG
GTTCCCGATCATGTTATTGCAGCAGCACTCACAGATGGTAAGCTGAGTTTACT
G***GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA***
GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAA
ATACTCTTAACTCCCCACACAAGGTGTATGCTCAAAGCGCTAACATTATTTTT
TCATACTCATCTCCGCTAGAAGGTTAACCATCTTATCCCACCGGGT

AB10R
CTTTTTGTGCCACTGCTGCTTTGGGTAAGATCCCTGTGGAGGTTCTGTGGTAG
GTCCCCATTAACTTATGAAGGACGGAAGAGGCAACTAAAAGTGTTCAGGTGG
TCAATCTCTCTGACTTGGCAGATTGTGTGTCATTGTGCCATGGATCAATGATC
ACACTTTCAGTCCCTGGTTTCCTGATCAGGACCTGATTATGTACTGCTGATGT
AATATTGATAAATAAATAAATAAATAAATAAAATCACACTTTCAGTCCCTGG
TTTCCTGATCAGGACCTGATTATGTACTGCTGATGTAATATT***GATAGATAGAT***
AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATT
AAATAAATAAATAAATAAATAAATAAATAAATAAATAAATATTCTATTGATA
TTCTGCCAGCAATAAATAAATAAATAAATAAATAAATATTCTATTGATATTCT
GCCAGCAATAAATAAATAAATAAATAAACAAACATTCTATTTATATTCTGGC
AGCCAGTACGTAACGGCGAACTGGAGACTGTATGCATTTATTGACAATGTTC
CAACTGCTGATTGATGCAATACTGNNCACCTTGTGAGCATTGTGGACATCTAC

CCTNGTTGAAGCNTGAG

AB11R = AB25R
ACATTTATACTGAATNNCGGAAGTAGCTGAAACGAATCTCAAACAAGTGNCG
CAGTTATCAAACCCTT***CTATCTATCTATCTATCTATCTATCTATCTATCTATCTA***
TCTATCTCCCTTTCTGCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCATATG
CACACACTAAGATAATGTTCAAAATTACGGT

AB13R
CTGGATCTCGGGTTCGAACACACGATCCTAGGTTTGTGGCATGCCTGAAGGA
CGCAAAACTGCATGGTGACGTTTCAGTACTGGTGATAGTCTCTTTATGAAGTT
GGTAGGAAATGGTTACATACTAACGATTACATATTGAAAGAGAGTGACAAAA
ACAACATATAGCAAAATGATGAATAAAGCCTGAAAACAGCAAGGTTGCGGT
AAGGTGGTATTGTGATTCAAAAGCTTA***GATAGATAGATAGATAGATAGATAGA***
TAGATAGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGA
GTGAGTGAGTGAGTGAGTGAGTGAGTGAATGAGTGAGTGAGTGAGTGAGTG
AGTGAGTGAGTGAGTGAGTATCTCAAAGGGATCGCCAAACTCGATTTTTCTTT
CAATGTTAGTACCTCAACAGATATGTTATAGTTGAAATGGCAACAGTATTTTT
ATGTTAAAAACAGCATATGAG

AB14R
CCATTAGCGCTGACATGTTAATGAGGTAGTGGGCTGTTATGGTTTTTTATTGT
CACTTCATTAGAGCTGTATTTTAGAGAAACATTCTCATCAGCTGTGAACTTTC
ATTTCAAGCTTGCATTAAGTTGAGGTTCAGGGCTGTTAAATGGCTTTCACTTC
AAGCTCACTAAGATAATAATTGCATGCCTCCTGTTTTTTT***CTATCTATCTATCT***
ATCTATCTATCTATCTATCTATCTATCTATCTATTAGTATGTTGAAATATGATAT
GACTTTTACAGCAGGAGAGCAACAAGCGTCAGTCTCTTCACTTAAACGAGAC
ACCCCGTGGCAGCTGGG

AB17R
CCAAGACAATAATGGGACGGTAGGGTAGCCAAATGATACAAGTGATTGTCAG
TTACGGCTAAGATCCGGGTACAATTTTTTAAATGGGTCCAATGTATAACGTCT
ATTTCTGGTCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCAC
TCACTCACTCACTCACTCA***CTATCTATCTATCTATCTATCTATCTATCTATCTAT***
CTATCTATCTATCTATCTATCTATCTATCTATCTATGGAGTCCACCAGAGCTAC
GTGGTATCAGATCCATAGCAGAAGGAAAAGCACCTTGCAAGATTGTACTATC
TTGACATGCGATATGTTGCCATCTGCAAGTACTTGAAACATGGCTATCTGAGC
AAATGTGATAACCACTGATGTATGATGAGGATATTTGTGTGACTGAACAGGG
ACATGTGCAGAAACAACTTGAAACTAACAACAGTTGCCCTGAGTAGCAGGTG
ACTGGGGG

AB21R = AB22R
CTATGTATTATTNAGCTCGGCCTCACGTACAATTAGCCTTCCTCCTACNNGTT
ACTCCATCAATTACCAATATCCATTGTTGTTTGCTATCCCTTTCATCCATACCA
TCATATTGTCTTTGTTAATCCCTTTCATTTATAATATCTTATTGACTTTGTTAAT

CCAAGTCATGTATATACGTCGGTTCCGTGCTGTACTACTACTGGCCTTCTAGG
TCGCATATAATCCCTCATAACCAACTAGTACCAAT***CTATCTATCTATCTATCTA***
TCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTA
TCTATCTATCTATCTATCTATCTATCTATCTATCTATCAAAACAACATAACCAA
AAAAACGAACAAACAAACAAAAACAAAAACGAAACATAACCAATGAAACCA
CCCCTTATTTCAGTAGTCCGAATTCGCCCTATAGTGAGTCGTATTACA

AB22R = AB21R
CTATGTATTATTTGCTCGGCCTCACGTACAATAGCCTTCCTCCTACTGTTACTC
CATCAATTACCAATATCCATTGTTGTTTGCTATCCCTTTCATCCATACCATCAT
ATTGTCTTTGTTAATCCCTTTCATCCATACCATCATATTGTCTTTGTTAATCCC
TTTCATTTATAATATCTTATTGACTTTGTTAATCCAAGTCATGTATATACGTCG
GTTCCGTGCTGTACTACTACTGGCCTTCTAGGTCGCTATTAATCCCTCATAAC
CAACTAGTAC***CTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC***
TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCAAC
AAAACAACATAACCCAAAAAACGAACAAACAAACAAAAACAAAAACGAAA
CATAACCAATGAAACCACCCCTTATTTCAG

AB23
TNNTNTNGTNCTAGTTATAATTNNGGNTTCCANNNTCCATTTCCATTTCCATT
ACCTTCCTCCCGTTNGCCCTTNGTCCATNNGCCAATCTCATATCAGCGTCAGT
TAATNGTNGGCTTCCCTNGTTAAGCTACCACCCTTGNGATCTTATTA***GATAGA***
TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA
GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAGAG
AGAGAGAGTCTTTATTTCCTCCAGTAATTACCGAAGATTACAAACAAAACATC
ATATTATCACAACATGT

AB24
ACCACCAAGGATGGCAATGCCTCAGCTACAGACAGGTGATTTTATATACATT
TTACATTCACATTCAAATGTTATAAATAAACTGTGCAAAAAAAAA***TAAATAAAT***
AAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAA
TAAATAAATAAATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA
GATAGACAGATAGAAGCTGGGTAGAAGGTTAAAAAGATTTCTTTTAACAGT

AB25 = AB11R
CTACNNNTTGTGTATACAGCTGGAAAACAATCAGTGGAGAAACCAACTTCTC
CTCATTCCCCTTAACGCAACAAACAACACAACTACAACCTACNGGGATACAG
CACGGTGCACTGACTGAGCGCAGTTTATTCATCAACATATATCTGCATTATCA
TCAACATCATCGGCATCACCACAGGATAATTTACATGAATGTAAACTAAATG
CTTCACATTTCCAACGTTTTATATACACAGGCATCGACGCCTATTCACGTATT
CTAATGGCTCGTCAGCGCATTCTCATCAGAGTATATCCCAAATCATATTCTCA
CAACTGGAGTGTTTGAGATTTTAACTATGGATACCGATCCATCTCACAACGTA
TTCTCTTGGCTCGACAGCACAACATGCAAACAACGCATAACATATTGACCAT
CATATGCATAGCAGCGGCTACTAAACACATTGAAAAAGAGAGAGATTCATCC
AGCCTGTTCTTCAATCGGCATCATACATAACTGAACATGTTTAATATAGCCAG

CCCCGAGTGCACCGAACCTGAGAATGAAGTAGTCCGAATTCGGACTAACCGT
AATTTTGAACATTATCTTAGTGTGTGCATAT***GAGAGAGAGAGAGAGAGAGAGA***
GAGAGAGAGAGAGGCAGAGAGGGA***GATAGATAGATAGATAGATAGATAGATA***
GATAGATAGATAGATAGANGGGTTTGATAACTGCGTCACTTGTNNGAGATTCG
TNTCAGCTACTTCCGCCATTCANTATAAATGT

AB30R
CTAAGGTGTATTTGCGCCCATCACGTTTTTTCAGTCCCGGATCCGAACATCTT
CTACATCATCCTGTCTCTTGTGAGCCAGTCTCCGGAGCAGCATCATACGCAGG
TGTTGAAACAATGGTGAAGTAATATTTCTTACCATTCTCTCCATACTTGCACT
GGGTAGCACTGCAGAGTGTAATGAGGAACACCACGTTCACGTCACCGACGTC
AACACCACGGCCACCACGTGGAAACAGCATGGTCGTTAACAGCGAACCATGT
CATGAACATCATTAGATACACAGAAACCCTCTTGATATCTGTCCAAACCACCC
ATGCGGCTGTCAGGTTATGGAAGAACTGTCTACAGGGATACTTCAAGCAAGG
AGTGATTGCGTGACGCCCGGCAAACGTGTCAGCTTGCATGATCTTTATATTAG
AAAGGAACGGAATGTGACAAGGGGGGATTTAACAATTGTATATACATGTGAT
GGAAGGAGTGGCAAAATCGGGAGTGGGGTCCCTGGGGTGATGATTGTGATGC
TTATTTATTTATTTATTTATTTATCTATCTATCTATCTATCTATCTATCTATCTAT
CTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT
CTATTTATTTATTTATTAAGGCAGTACAGTCTTCGCCGAGTGATTATAG

AB31
CTACGTAGCGGGCGGGTAACTTATAGCCTTGATCATTGATTTTGGAAGTTGTT
TGATGTTGATGGGGTTTTTTTCAGTTTTCTGACATTGTCTCATACCTGTTTGTA
TCGGTTTAGTCTCTATTCGCCTACATCTCAATGTAAACTTGTTTTGATATCTTT
TTCCTTTCTTACTACCTTTTAACCTTATTTTATATCGTGTGTTTTAGTCATGTGT
ATATAGGGTAATCATTTTATGTCTCTCATAACTGTTTTAGAGTGGCTCATTAC
ATATTTTATTGTCATTGTTTTCCTGTATATATTTTAACTATGTAATGAGGTGAG
ACGTTAATAAA***CTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT***
CTATCTCAATTCGCCTTCAACCCATATCGCATACCAAAACAATGTAGGCTAAT
TGGGATTAGAAAATGCACCCCCATACTCTTATAATGAGCCTGACTCACTGGTC
AGGGCCATACTAATCTTGATCGACTG

APPENDIX 7: *HALIOTIS IRIS* MICROSATELLITE ALLELE FREQUENCIES

AB14 Allele	AHU 22	CBL 14	CCB 19	CRW 19	DBL 15	DSD 14	EAI 22	GLN 20	GOB 19	IHM 20	JCH 19	MAT 18	MTB 21	NPT 19	OCH 13	OLB 17	OPT 19	PHD 17	SPB 21	STR 14	TCL 19	TIM 22	TSK 18	WLG 19	WST 15
169	0.068	–	–	–	–	–	0.045	0.025	–	–	–	–	0.048	–	–	–	–	–	0.024	0.036	–	–	–	–	–
173	0.023	0.036	0.105	0.132	–	–	–	0.050	0.105	0.025	0.132	0.028	0.024	–	–	0.059	–	–	0.071	0.071	–	0.068	0.028	0.026	0.167
174	–	–	–	–	–	–	–	–	–	–	0.026	–	–	–	–	–	–	–	–	–	–	–	–	–	–
177	0.182	0.143	0.237	0.211	0.133	0.143	0.114	0.225	0.158	0.250	0.053	0.139	0.095	0.211	0.231	0.176	0.158	0.235	0.167	0.143	0.079	0.136	0.222	0.105	0.033
178	–	–	–	–	–	–	–	–	0.026	0.025	0.026	–	–	–	–	–	–	–	–	–	–	0.023	–	–	–
181	0.136	0.071	0.079	0.158	0.200	0.107	0.318	0.200	0.079	0.175	0.079	0.167	0.143	0.132	0.346	0.088	0.263	0.235	0.167	–	0.263	0.068	0.083	0.184	0.167
182	–	–	–	–	–	–	–	0.025	–	–	–	–	–	–	–	–	–	–	–	0.036	–	–	–	–	–
185	0.159	0.321	0.079	0.079	0.233	0.036	0.205	0.150	0.105	0.100	0.237	0.306	0.095	0.105	0.308	0.176	0.211	0.206	0.143	0.214	0.184	0.068	0.056	0.263	0.167
186	–	–	–	–	–	–	–	0.025	–	–	0.026	–	–	–	–	–	–	–	–	–	–	–	–	–	–
189	0.159	0.143	0.105	0.158	0.200	0.286	0.136	0.175	0.079	0.100	0.132	0.111	0.214	0.079	0.077	0.118	0.132	0.059	0.167	0.107	0.105	0.250	0.194	0.184	0.133
190	0.023	–	–	–	–	–	–	–	–	–	0.026	–	–	–	–	–	–	–	–	–	–	–	–	–	–
193	0.136	0.071	0.132	0.184	0.067	0.143	0.091	0.075	0.132	0.075	0.105	0.194	0.190	0.132	–	0.118	0.079	0.147	0.167	0.036	0.079	0.159	0.167	0.132	0.100
194	–	–	–	–	–	–	–	0.050	–	–	0.026	–	–	–	–	–	–	–	–	–	–	0.023	–	–	–
196	–	–	–	–	–	–	–	–	–	0.050	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
197	0.023	–	0.026	0.026	0.100	0.036	–	–	0.079	0.100	–	–	0.071	0.079	–	0.059	0.053	–	0.048	0.071	0.079	0.068	0.028	0.026	–
200	0.045	0.071	0.158	0.026	0.067	0.214	0.045	–	0.158	0.050	0.105	–	0.095	0.237	–	0.088	0.053	0.059	–	0.107	0.184	0.114	0.139	0.026	0.200
203	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.036	–	–	0.056	–	–
204	0.023	0.143	0.079	–	–	0.036	0.023	–	0.079	0.050	–	0.056	0.024	–	0.038	0.118	0.026	0.029	0.024	0.143	0.026	0.023	0.028	0.026	0.033
208	–	–	–	0.026	–	–	0.023	–	–	–	–	–	–	0.026	–	–	0.026	–	–	–	–	–	–	–	–
212	0.023	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
216	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.026	–
220	–	–	–	–	–	–	–	–	–	–	0.026	–	–	–	–	–	–	0.029	–	–	–	–	–	–	–
224	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.024	–	–	–	–	–	–
AB21 Allele	AHU 22	CBL 14	CCB 17	CRW 19	DBL 15	DSD 14	EAI 18	GLN 20	GOB 19	IHM 20	JCH 20	MAT 18	MTB 22	NPT 20	OCH 13	OLB 19	OPT 18	PHD 16	SPB 20	STR 13	TCL 20	TIM 22	TSK 13	WLG 20	WST 15
167	–	0.036	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.028	–	0.025	–	–	–	–	–	–
171	–	–	–	–	–	–	–	0.025	–	0.025	–	0.028	–	–	–	–	–	–	0.050	–	–	–	–	–	–
175	0.023	–	–	–	–	–	–	0.025	–	–	–	0.028	0.023	–	–	–	–	0.031	–	–	–	–	–	–	–
179	0.023	–	–	–	0.033	–	–	–	–	–	–	–	–	–	–	–	–	–	0.025	–	0.025	–	–	–	–
180	–	–	0.059	0.053	–	–	–	–	–	–	–	–	–	–	0.038	–	–	–	0.025	0.038	0.050	0.023	0.038	0.050	–
181	–	0.036	–	0.026	–	–	–	–	0.053	–	–	–	–	0.025	0.038	–	–	–	–	–	0.025	–	0.038	–	–
182	0.023	–	–	–	–	–	–	–	–	–	–	0.028	–	–	–	–	–	–	–	–	–	–	–	0.050	–
183	0.023	–	–	–	–	–	0.056	0.025	–	–	–	0.056	0.023	–	–	–	0.028	–	0.025	–	0.050	–	–	–	–
184	–	–	–	–	0.033	–	–	0.025	–	–	–	0.028	0.023	0.050	–	–	–	0.031	0.025	–	0.025	–	–	–	–
186	–	–	–	0.026	–	–	–	–	–	0.025	–	0.028	–	–	–	–	0.028	–	–	–	–	–	–	–	–
187	0.045	0.143	0.029	0.026	0.067	–	0.056	0.025	0.105	0.025	0.025	0.028	0.068	0.075	–	0.079	0.111	0.031	–	0.038	0.025	0.068	–	0.025	0.067
189	–	–	–	–	–	–	–	–	–	–	0.050	–	–	–	–	–	–	–	–	–	–	–	–	–	–
190	0.045	–	0.059	–	–	–	0.028	–	–	–	–	–	–	–	0.038	–	0.028	–	0.025	–	0.025	0.045	–	–	–
191	0.091	0.036	0.059	0.079	0.133	–	0.278	0.025	0.026	0.025	0.050	0.056	0.023	–	–	–	0.083	0.031	0.075	0.038	0.075	0.091	0.077	0.050	–
192	–	0.036	–	–	–	–	–	–	0.026	–	–	0.028	0.023	0.075	0.038	0.026	–	0.031	–	–	–	–	–	–	–

193	–	–	–	–	–	–	–	–	–	0.025	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
194	0.023	–	–	–	0.067	–	–	0.025	–	0.025	–	–	–	–	–	–	–	–	–	–	–	–	–	0.025	–
195	0.023	–	–	0.026	0.133	–	0.028	0.050	–	0.050	–	–	–	–	0.038	0.026	0.028	0.063	0.050	–	0.025	–	–	0.050	–
198	–	–	0.029	0.079	–	0.036	–	0.025	0.026	–	–	–	0.023	–	–	–	0.028	–	0.075	0.038	0.075	0.045	0.038	0.025	0.033
199	–	–	–	–	–	–	0.028	0.025	–	–	–	–	–	–	–	–	0.028	–	0.075	–	–	–	–	–	–
200	0.068	–	–	–	–	–	–	0.025	–	–	0.025	–	–	–	–	–	–	–	–	0.038	–	–	–	–	–
201	0.023	–	–	–	–	–	–	0.050	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
202	0.045	0.107	–	0.053	0.033	0.071	0.056	0.025	0.026	0.125	0.050	–	0.045	0.075	0.038	0.053	0.028	0.031	0.025	–	0.075	0.045	0.077	–	0.100
203	0.091	–	–	0.026	–	–	0.083	–	–	–	–	0.056	–	–	–	–	0.028	–	–	–	–	–	–	0.075	–
205	–	–	–	–	0.033	–	–	–	–	0.025	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
206	–	–	0.029	0.079	–	0.036	–	0.025	–	0.025	–	0.028	–	–	–	0.053	0.028	0.031	–	–	–	–	–	–	–
207	0.023	–	0.029	–	0.033	–	0.028	–	–	0.050	–	–	–	0.025	0.077	–	0.028	0.063	0.050	–	–	–	–	–	–
208	–	–	–	–	–	–	–	0.025	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
209	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.038	–	–	–	–	–	–	–	–	–	–
210	0.045	0.036	–	0.026	–	0.036	–	–	–	0.025	–	0.028	0.023	0.050	0.038	–	–	–	0.025	–	–	–	–	0.050	–
211	0.045	–	–	–	–	–	0.028	–	–	–	–	–	–	–	–	–	0.056	–	–	–	–	–	–	–	–
212	0.045	–	–	0.026	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.025	–	–	0.025	–
213	–	–	–	–	–	–	–	–	–	0.050	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
214	0.045	0.036	0.118	0.026	–	0.143	0.083	0.050	0.132	–	0.075	0.056	0.091	0.050	–	0.026	0.083	0.063	0.025	0.038	0.025	0.091	0.077	0.025	0.100
215	0.023	–	–	–	–	–	–	–	–	–	–	–	–	–	0.038	–	–	–	–	–	–	–	–	–	–
216	0.023	–	–	–	–	–	0.028	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
218	–	0.071	0.059	–	–	0.036	0.028	0.050	0.026	–	0.025	0.028	0.045	0.025	–	0.079	0.028	0.031	–	–	0.025	0.091	0.038	0.025	0.100
219	–	–	–	0.026	–	–	–	–	–	–	–	–	–	–	0.038	–	–	–	–	–	–	–	–	–	–
220	–	–	–	0.026	0.033	–	–	0.025	–	–	–	–	–	–	0.038	–	–	–	–	–	–	–	–	0.025	–
222	0.068	0.036	0.029	–	0.033	0.071	–	0.025	0.053	–	0.025	–	0.045	0.075	–	0.079	–	0.094	0.025	0.038	0.025	–	0.077	0.150	–
223	–	–	–	0.053	0.033	–	0.028	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
224	–	0.036	–	0.026	–	–	–	–	–	0.025	–	–	0.023	–	0.038	–	–	–	–	0.038	–	–	–	–	–
226	–	0.036	–	–	–	–	0.028	0.050	–	0.050	0.050	0.028	0.023	–	–	–	–	–	0.025	–	–	–	–	–	–
227	–	–	–	–	–	–	–	–	–	0.025	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
228	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.028	–	–	0.038	–	–	–	–	–
230	–	–	0.029	–	–	0.036	–	0.050	–	0.025	0.050	0.111	0.045	–	–	0.026	0.056	0.031	0.075	0.077	–	0.023	0.077	0.025	–
231	–	–	0.029	0.026	–	–	–	0.025	–	0.025	–	–	–	0.025	–	0.026	–	0.031	–	–	–	–	–	–	–
232	–	–	–	–	–	–	–	0.050	–	0.050	–	–	–	–	–	–	–	–	0.025	–	–	–	–	–	–
233	–	–	–	–	–	–	–	–	0.026	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
234	–	0.036	0.029	–	0.033	0.071	0.028	–	–	0.075	0.025	–	0.023	0.075	–	0.105	–	0.031	0.025	0.077	0.075	0.023	–	0.050	0.100
235	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.050	–	–	–	–
236	0.023	–	–	0.026	–	0.036	–	–	–	–	–	–	–	–	0.077	–	–	–	–	–	–	–	–	–	–
238	0.023	0.071	0.059	–	0.033	0.036	0.028	0.050	0.079	0.025	0.050	0.056	–	0.050	–	0.105	0.111	0.094	0.025	–	–	0.045	–	–	0.067
239	–	–	0.029	–	0.033	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.023	–	–	–
240	0.023	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
242	–	–	0.059	–	0.067	–	0.028	0.100	0.053	0.050	0.125	0.083	0.114	–	0.038	0.079	0.028	0.063	0.050	0.115	0.025	0.091	0.038	0.050	0.100
243	–	0.071	–	0.053	0.033	–	–	–	0.053	–	–	–	–	–	0.038	0.079	–	0.063	–	0.038	–	0.045	–	0.025	–	

244	–	–	–	–	–	–	–	0.050	–	–	0.025	–	–	–	–	–	–	–	–	–	–	–	–	–	–
245	–	–	–	–	–	–	–	–	–	–	0.050	–	–	–	–	–	–	–	–	–	–	–	–	–	–
246	0.023	–	0.088	0.026	–	0.143	–	–	0.026	0.050	0.150	0.028	–	0.050	0.038	–	0.056	0.031	0.025	0.038	0.075	0.045	0.038	0.025	0.067
247	–	–	0.029	–	0.033	0.036	–	–	–	–	–	0.028	–	0.025	0.038	0.026	–	–	–	–	–	–	–	–	–
248	–	–	–	0.026	–	–	–	–	0.053	–	–	0.028	0.091	–	–	–	–	0.031	–	–	–	–	–	–	–
249	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.045	–	–	–
250	0.045	0.071	0.029	–	–	0.071	0.028	–	–	0.050	–	0.028	–	0.025	–	–	0.056	–	0.025	0.192	–	–	0.115	0.025	0.067
251	–	–	–	0.026	–	0.036	–	–	0.079	–	–	–	–	0.025	0.038	0.026	–	–	–	–	–	–	–	–	0.033
252	–	–	–	–	–	–	–	–	–	–	0.025	0.028	0.023	–	0.038	–	–	0.031	–	0.038	–	–	–	–	–
254	–	–	0.029	–	–	–	0.028	–	–	–	0.050	0.028	–	–	–	–	–	–	0.100	0.077	0.075	0.114	0.038	–	0.067
255	–	–	–	0.026	0.033	0.036	–	–	0.026	–	–	0.028	0.023	0.050	–	0.026	–	0.031	–	–	–	–	–	0.025	–
256	–	–	–	–	–	–	–	0.025	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.025	–
258	–	–	0.059	–	–	–	–	–	–	0.050	–	–	–	–	–	–	–	–	–	–	–	–	0.038	0.050	0.100
259	–	–	–	0.026	–	0.036	–	–	0.079	–	–	–	0.068	0.050	0.115	–	–	–	–	–	–	0.023	0.038	–	–
260	–	–	–	–	–	–	–	–	–	–	–	0.028	0.045	–	–	–	–	–	–	–	–	–	0.038	–	–
262	–	0.036	0.029	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.100	–	0.038	0.025	–
263	–	–	–	0.026	0.033	–	–	–	–	–	0.025	–	0.045	–	0.038	0.026	–	–	–	–	–	–	–	–	–
264	–	–	–	–	–	–	–	–	–	–	0.025	–	–	0.025	–	–	–	–	–	–	–	–	–	–	–
267	–	–	–	0.026	–	–	–	–	–	–	0.025	–	–	0.050	–	0.026	–	0.031	–	–	0.025	0.023	0.077	–	–
268	–	–	–	–	–	–	–	–	–	–	–	–	0.023	–	–	–	–	–	–	–	–	–	–	–	–
271	–	–	–	–	0.033	0.036	–	–	0.053	–	–	–	–	–	–	0.026	–	–	–	–	–	–	–	–	–
273	–	–	–	–	–	–	–	–	–	–	–	–	–	0.025	–	–	–	–	–	–	–	–	–	–	–
275	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.025	–
276	–	–	–	0.026	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
279	–	–	–	–	–	–	–	0.025	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
280	–	0.036	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
287	–	0.036	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
AB31 Allele	22	14	19	18	14	14	23	20	19	20	20	14	22	19	13	19	19	17	21	15	20	23	19	20	15
162	–	–	0.053	0.083	0.071	0.036	0.043	–	0.026	–	0.050	–	–	–	–	0.026	–	–	–	–	0.050	0.022	0.026	0.075	0.033
170	0.023	–	–	–	–	–	0.022	–	–	–	–	–	–	–	–	–	0.053	–	–	–	–	–	–	–	–
174	0.068	–	0.079	0.056	0.036	–	0.022	0.050	–	0.025	–	0.036	0.023	0.026	–	–	0.026	–	–	–	0.050	–	0.026	–	–
175	–	–	–	–	–	–	–	0.025	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
178	0.114	0.036	0.079	0.139	0.107	0.036	0.109	0.125	0.026	0.175	0.075	0.071	0.045	0.053	0.115	0.053	0.053	0.029	0.095	0.100	0.150	0.022	0.053	0.125	0.133
181	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.029	–	–	–	–	–	–	–
182	0.159	0.107	0.079	0.028	0.071	0.143	0.065	0.175	0.079	0.125	0.075	0.179	0.091	–	0.077	0.132	0.158	0.147	0.119	0.067	0.150	0.130	0.026	0.025	0.067
183	–	–	–	–	–	–	–	–	–	–	0.050	–	–	–	–	–	–	–	–	0.033	–	–	–	–	–
186	0.068	0.393	0.316	0.194	0.107	0.286	0.283	0.200	0.316	0.250	0.325	0.357	0.341	0.500	0.115	0.395	0.263	0.471	0.262	0.433	0.200	0.391	0.474	0.325	0.333
187	–	–	–	–	–	–	–	0.025	–	–	–	–	–	–	–	–	–	–	–	0.033	–	–	0.026	–	–
189	0.045	–	–	–	–	0.036	–	–	–	0.025	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
190	0.045	–	0.079	0.083	0.214	0.107	0.087	0.150	0.132	0.100	0.075	0.071	0.114	0.132	0.077	0.132	0.105	0.118	0.119	0.067	0.050	0.109	0.105	0.125	0.033
191	0.045	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–

192	–	–	–	–	–	–	–	–	–	–	0.025	–	–	–	–	–	–	–	–	–	–	–	–	–	0.033
194	0.091	0.179	0.184	0.139	0.143	0.143	0.196	0.025	0.158	0.100	0.050	0.143	0.136	0.079	0.154	0.105	0.079	0.059	0.071	0.167	0.225	0.087	0.026	0.200	0.133
195	0.023	–	–	–	–	–	–	–	–	–	0.025	–	–	–	–	–	–	–	–	–	–	–	–	–	–
198	0.068	0.179	0.132	0.111	0.071	0.107	0.065	0.050	0.079	–	0.125	0.107	0.159	0.158	0.154	0.105	0.079	0.118	0.048	0.067	0.025	0.152	0.158	0.050	0.167
200	–	–	–	–	–	–	–	–	–	–	0.075	–	–	–	–	–	–	–	–	–	–	0.043	0.026	–	–
201	–	–	–	–	–	–	–	–	0.053	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
202	0.023	–	–	0.111	0.036	0.036	0.065	0.050	0.053	0.025	–	0.036	0.023	0.026	0.115	–	0.026	–	0.048	–	0.025	–	0.026	0.025	–
206	0.023	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.026	–	–	0.024	–	–	–	–	–	–
210	–	–	–	–	–	–	–	–	0.026	0.025	0.025	–	–	–	–	–	–	–	–	–	–	–	–	–	–
214	0.023	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.024	–	–	–	–	–	–
217	–	–	–	–	0.036	–	–	–	–	0.050	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
221	–	–	–	–	–	–	–	0.025	–	–	–	–	–	–	–	–	0.026	–	–	–	–	–	–	–	–
225	–	–	–	–	0.036	–	–	0.025	–	–	–	–	–	–	–	–	–	–	0.024	–	–	–	–	0.025	–
227	0.023	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
229	0.023	0.036	–	–	0.036	–	–	0.025	–	0.025	–	–	0.023	–	0.038	–	–	–	0.048	–	0.025	–	–	–	–
230	0.023	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
233	0.023	–	–	0.028	0.036	–	–	0.025	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
237	–	–	–	–	–	–	–	–	–	–	–	–	0.045	–	0.038	–	0.026	–	–	–	–	–	–	–	–
239	0.023	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
241	–	–	–	0.028	–	–	–	–	–	–	–	–	–	–	–	–	0.053	–	0.048	–	0.025	–	–	–	–
245	0.045	–	–	–	–	–	–	–	–	0.025	–	–	–	–	0.038	0.026	0.026	–	–	–	–	–	0.026	0.025	–
249	–	–	–	–	–	–	–	0.025	0.026	0.050	–	–	–	–	–	–	0.026	–	0.024	–	–	–	–	–	–
253	0.023	–	–	–	–	–	0.043	–	0.026	–	–	–	–	–	–	–	–	–	0.024	–	–	–	–	–	0.033
257	–	0.036	–	–	–	0.036	–	–	–	–	–	–	–	0.026	–	–	–	–	–	–	–	–	–	–	–
261	–	0.036	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.029	–	0.033	–	–	–	–	–
265	–	–	–	–	–	0.036	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.043	–	–	–
269	–	–	–	–	–	–	–	–	–	–	0.025	–	–	–	0.038	–	–	–	0.024	–	–	–	–	–	–
273	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.038	–	–	–	–	–	–	–	–	–	–
277	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.025	–	–	–	0.033

APPENDIX 8: LYSIN HAPLOTYPES

Listed below is the alignment in FASTA format for the 24 lysin haplotypes. Labels refer to the numbered haplotypes in Figure 5.4.

>h1
GCGTAGGTGCCGCAATTGGAAAACGCATTCCCTTGGAAATAACGTACAGCTTTTTAGTCAGGAGGAACTTGATACCTAAA
TGGAAACCGTACATGGCACGACTTATGGCCAGTAAGTATTGCATATTATTAAGAACTGATCTCAATATTTTGCGTTGATG
GAGTTATAAACATACATCCGGATGCGCAGAATTGGGATAAATCCGAAATTACGTTGTAAAAAGAATAACTATACTTCACA
GATTATTTCATCAGTAACATAAAAACTGTACACATCGTTCACCGCAATAATGAGGCACAAACCACGCCCACTGTATTGTC
AGCCTGTAAGTAACACAGTGGTCCCTAGTTGGCAAGGCCTTAGACGGTCACCAGGGAACGAAATCAGTTTTGTTTCCTGT
TAAGGTATCCCGAAAGCAAATAGAATGTCAAACATCTATGGTTTAATGCTAATTTGCTAAGTTATCTCTATTTGGACTAG
TAGTGTGTCTCTGAAATCACTTTTATTTTATTTATCTATATATTTATTAAGAGCTAAGAAAATGGAACTGGTCGTCATGC
CTGGTGACTTGCCTGATTGCCCCGCAGACATGGATCGTTGTAAAAATTATAAGCTGACTCCAGACTTATTTACATGTCGC
CTTCATATAGCTGGAACTTTTGGTGAATGTCGCGTAACAAAACACTTTATTCGTAGTTTTAAGAGGTCATCCCTTCATTT
GAAGTGTATTGTTTGTTTTTTAAAAATTTCTAATGCACGTTTGTTTTGTTTCCACAGGACGCC

>h2
GCGTAGGTGCCGCAATTGGAAAACGCATTCCCTTGGAAATAACGTACAGCTTTTTAGTCAGGAGGAACTTGATACCTAAA
TGGAAACCGTACATGGCACGACTTATGGCCAGTAAGTATTGCATATTATTAAGAACTGATCTCAATATTTTGCGTTGATG
GAGTTATAAACATACATCCGGATGCGCAGAATTGGGATAAATCCGAAATTACGTTGTAAAAAGAATAACTATACTTCACA
GATTATTTCATCAGTAACATAAAAACTGTACACATCGTTCACCGCAATAATGAGGCACAAACCACGCCCACTGTATTGTC
AGCCTGTAAGTAACACAGTGGTCCCTAGTTGGCAAGGCCTTAGACGGTCACCAGGGAACGAAATCAGTTTTGTTTCCTGT
TAAGGTATCCCGAAAGCAAATAGAATGTCAAACATCTATGGTTTAATGCTAATTTGCTAAGTTATCTCTATTTGGACTAG
TAGTGTGTCTCTGAAATCACTTTTATTTTATTTATCTATATATTTATTAAGAGCTAAGAAAATGGAACTGGTCGTCATGC
CTGGTGACTTGCCTGATTGCCCCGCAGACATGGATCGTTGTAAAAATTATAAGCTGACTCCAGACTTATTTACATGTCGC
CTTCATATAGCTGGAACTTTTGGTGAATGTCGCGTAACAAAACACTTTATTCGTAGTTTTAAGAGGTCATCCCTTCATTT
GAAGTGTATTGTTTGTTTTTTAAAAATTTCTAATGCACGTTTGTTTTATTTCCACAGGACGCC

>h3
GCGTAGGTGCCGCAATTGGAAAACGCATTCCCTTGGAAATAACGTACAGCTTTTTAGTCAGGAGGAACTTGATACCTAAA
TGGAAACCGTACATGGCACGACTTATGGCCAGTAAGTATTGCATATTATTAAGAACTGATCTCAATATTTTGCGTTGATG
GAGTTATAAACATACATCCGGATGCGCAGAATTGGGATAAATCCGAAATTACGTTGTAAAAAGAATAACTATACTTCACA
GATTATTTCATCAGTAACATAAAAACTGTACACATCGTTCACCGCAATAATGAGGCACAAACCACGCCCACTGTATTGTC
AGCCTGTAAGTAACACAGTGGTCCCTAGTTGGCAAGGCCTTAGACGGTCACCAGGGAACGAAATCAGTTTTGTTTCCTGT
TAAGGTATCCCGAAAGCAAATAGAATGTCAAACATCTATGGTTTAATGCTAATTTGCTAAGTTATCTCTATTTGGACTAG
TAGTGTGTCTCTGAAATCACTTTTATTTTATTTATCTATATATTTATTAAGAGCTAAGAAAATGGAACTGGTCGTCATGC
CTGGTGACTTGCCTGATTGCCCCGCAGACATGGATCGTTGTAAAAATTATAAGCTGACTCCAGACTTATTTACATGTCGC
CTTCATATAGCTGGAACTTTTGGTGAATGTCGCGTAACAAAACACTTTATTCGTAGTTTTAAGAGGTCATCCCTTCATTT
GAAGTGTATTGTTTGTTTTTTAAAAATATCTAATGCACGTTTGTTTTGTTTCCACAGGACGCC

>h4
GCGTAGGTGCCGCAATTGGAAAACGCATTCCCTTGGAAATAACGTACAGCTTTTTAGTCAGGAGGAACTTGATACCTAAA
TGGAAACCGTACATGGCACGACTTATGGCCAGTAAGTATTGCATATTATTAAGAACTGATCTCAATATTTTGCGTTGATG
GAGTTATAAACATACATCCGGATGCGCAGAATTGGGATAAATCCGAAATTACGTTGTAAAAAGAATAACTATACTTCACA
GATTATTTCATCAGTAACATAAAAACTGTACACATCGTTCACCGCAATAATGAGGCACAAACCACGCCCACTGTATTGTC
AGCCTGTAAGTAACACAGTGGTCCCTAGTTGGCAAGGCCTTAGACGGTCACCAGGGAACGAAATCAGTTTTGTTTCCTGT
TAAGGTATCCCGAAAGCAAATAGAATGTCAAACATCTATGGTTTAATGCTAATTTGCTAAGTTATCTCTATTTGGACTAG
TAGTGTGTCTCTGAAATCACTTTTATTTTATTTATCTATATATTTATTAAGAGCTAAGAAAATGGAACTGGTCGTCATGC
CTGGTGACTTGCCTGATTGCCCCGCAGACATGGATCGTTGTAAAAATTATAAGCTGACTCCAGACTTATTTACATGTCGC
CTTCATATAGCTGGAACTTTTGGTGAATGTCGCGTAACAAAACACTTTATTCGTAGTTTTAAGAGGTCATCCCTTCATTT
GAAGTGTATTGTTTGTTTTTTTAAAAATTTCTAATGCACGTTTGTTTTGTTTCCACAGGACGCC

>h5
GCGTAGGTGCCGCAATTGGAAAACGCATTCCCTTGGAAATAACGTACAGCTTTTTAGTCAGGAGGAACTTGATACCTAAA

TGGAAACCGTACATGGCACGACTTATGGCCAGTAAGTATTGCATATTATTAAGAACTGATCTCAATATTTTGCGTTGATG
GAGTTATAAACATACATCCGGATGCGCAGAATTGGGATAAATCCGAAATTACGTTGTAAAAAGAATAACTATACTTCACA
GATTATTTCATCAGTAACATAAAAACTGTACACATCGTTCACCGCAATAATGAGGCACAAACCACGCCCACTGTATTGTC
AGCCTGTAAGTAACACAGTGGTCCCTAGTTGGCAAGGCCTTAGACGGTCACCAGGGAACGAAATCAGTTTTGTTTCCTGT
TAAGGTATCCCGAAAGCAAATAGAATGTCAAACATCTATGGTTTAATGCTAATTTGCTAAGTTATCTCTATTTGGACTAG
TAGTGTGTCTCTGAAATCACTTTTATTTTATTTATCTATATATTTATTAAGAGCTAAGAAAATGGAACTGGTCGTCATGC
CTGGTGACTTGCCTGATTGCCCCGCAGACATGGATCGTTGTAAAAATTATAAGCTGACTCCAGACTTATTTACATGTCGC
CTTCATATAGCTGGAACTTTTGGTGAATGTCGCGTAACAAAACACTTTATTTGTAGTTTTAAGAGGTCATCCCTTCATTT
GAAGTGTATTGTTTGTTTTTTAAAAATTTCTAATGCACGTTTGTTTTGTTTCCACAGGACGCC

>h6
GCGTAGGTGCCGCAATTGGAAAACGCATTCCCTTGGAAATAACGTACAGCTTTTTAGTCAGGAGGAACTTGATACCTAAA
TGGAAACCGTACATGGCACGACTTATGGCCAGTAAGTATTGCATATTATTAAGAACTGATCTCAATATTTTGCGTTGATG
GAGTTATAAACATACATCCGGATGCGCAGAATTGGGATAAATCCGAAATTACGTTGTAAAAAGAATAACTATACTTCACA
GATTATTTCATCAGTAACATAAAAACTGTACACATCGTTCACCGCAATAATGAGGCACAAACCACGCCCACTGTATTGTC
AGCCTGTAAGTAACACAGTGGTCCCTAGTTGGCAAGGCCTTAGACGGTCACCAGGGAACGAAATCAGTTTTGTTTCCTGT
TAAGGTATCCCGAAAGCAAATAGAATGTCAAACATCTATGGTTTAATGCTAATTTGCTAAGTTATCTCTATTTGGACTAG
TAGTGTGTCTCTGAAATCACTTTTATTTTATTTATCTATATATTTATTAAGAGCTAAGAAAATGGAACTGGTCGTCATGC
CTGGTGACTTGCCTGATTGCCCCGCAGACATGGATCGTTGTAAAAATTATAAGCTGACTCCAGACTTATTTACATGTCGC
CTTCATATAGCTGGAACTTTTGGTGAATGTCGCGTAACAAAACACATTATTCGTAGTTTTAAGAGGTCATCCCTTCATTT
GAAGTGTATTGTTTGTTTTTTAAAAATTTCTAATGCACGTTTGTTTTGTTTCCACAGGACGCC

>h7
GCGTAGGTGCCGCAATTGGAAAACGCATTCCCTTGGAAATAACGTACAGCTTTTTAGTCAGGAGGAACTTGATACCTAAA
TGGAAACCGTACATGGCACGACTTATGGCCAGTAAGTATTGCATATTATTAAGAACTGATCTCAATATTTTGCGTTGATG
GAGTTATAAACATACATCCGGATGCGCAGAATTGGGATAAATCCGAAATTACGTTGTAAAAAGAATAACTATACTTCACA
GATTATTTCATCAGTAACATAAAAACTGTACACATCGTTCACCGCAATAATGAGGCACAAACCACGCCCACTGTATTGTC
AGCCTGTAAGTAACACAGTGGTCCCTAGTTGGCAAGGCCTTAGACGGTCACCAGGGAACGAAATCAGTTTTGTTTCCTGT
TAAGGTATCCCGAAAGCAAATAGAATGTCAAACATCTATGGTTTAATGCTAATTTGCTAAGTTATCTCTATTTGGACTAG
TAGTGTGTCTCTGAAATCACTTTTATTTTATTTATCTATATATTTATTAAGAGCTAAGAAAATGGAACTGGTCGTCATGC
CTGGTGACTTGCCTGATTGCCCCGCAGACATGGATCGTTGTAAAAATTATAAGCTGACTCCAGACTTATTTACATGTCGC
CTTCATATAGCTGGAACTTTGGGTGAATGTCGCGTAACAAAACACTTTATTCGTAGTTTTAAGAGGTCATCCCTTCATTT
GAAGTGTATTGTTTGTTTTTTAAAAATTTCTAATGCACGTTTGTTTTGTTTCCACAGGACGCC

>h8
GCGTAGGTGCCGCAATTGGAAAACGCATTCCCTTGGAAATAACGTACAGCTTTTTAGTCAGGAGGAACTTGATACCTAAA
TGGAAACCGTACATGGCACGACTTATGGCCAGTAAGTATTGCATATTATTAAGAACTGATCTCAATATTTTGCGTTGATG
GAGTTATAAACATACATCCGGATGCGCAGAATTGGGATAAATCCGAAATTACGTTGTAAAAAGAATAACTATACTTCACA
GATTATTTCATCAGTAACATAAAAACTGTACACATCGTTCACCGCAATAATGAGGCACAAACCACGCCCACTGTATTGTC
AGCCTGTAAGTAACACAGTGGTCCCTAGTTGGCAAGGCCTTAGACGGTCACCAGGGAACGAAATCAGTTTTGTTTCCTGT
TAAGGTATCCCGAAAGCAAATAGAATGTCAAACATCTATGGTTTAATGCTAATTTGCTAAGTTATCTCTATTTGGACTAG
TAGTGTGTCTCTGATATCACTTTTATTTTATTTATCTATATATTTATTAAGAGCTAAGAAAATGGAACTGGTCGTCATGC
CTGGTGACTTGCCTGATTGCCCCGCAGACATGGATCGTTGTAAAAATTATAAGCTGACTCCAGACTTATTTACATGTCGC
CTTCATATAGCTGGAACTTTTGGTGAATGTCGCGTAACAAAACACTTTATTCGTAGTTTTAAGAGGTCATCCCTTCATTT
GAAGTGTATTGTTTGTTTTTTAAAAATTTCTAATGCACGTTTGTTTTGTTTCCACAGGACGCC

>h9
GCGTAGGTGCCGCAATTGGAAAACGCATTCCCTTGGAAATAACGTACAGCTTTTTAGTCAGGAGGAACTTGATACCTAAA
TGGAAACCGTACATGGCACGACTTATGGCCAGTAAGTATTGCATATTATTAAGAACTGATCTCAATATTTTGCGTTGATG
GAGTTATAAACATACATCCGGATGCGCAGAATTGGGATAAATCCGAAATTACGTTGTAAAAAGAATAACTATACTTCACA
GATTATTTCATCAGTAACATAAAAACTGTACACATCGTTCACCGCAATAATGAGGCACAAACCACGCCCACTGTATTGTC
AGCCTGTAAGTAACACAGTGGTCCCTAGTTGGCAAGGCCTTAGACGGTCACCAGGGAACGAAATCAGTTTTGTTTCCTGT
TAAGGTATCCCGAAAGCAAATAGAATGTCAAACATTTATGGTTTAATGCTAATTTGCTAAGTTATCTCTATTTGGACTAG
TAGTGTGTCTCTGAAATCACTTTTATTTTATTTATCTATATATTTATTAAGAGCTAAGAAAATGGAACTGGTCGTCATGC
CTGGTGACTTGCCTGATTGCCCCGCAGACATGGATCGTTGTAAAAATTATAAGCTGACTCCAGACTTATTTACATGTCGC
CTTCATATAGCTGGAACTTTTGGTGAATGTCGCGTAACAAAACACTTTATTCGTAGTTTTAAGAGGTCATCCCTTCATTT
GAAGTGTATTGTTTGTTTTTTAAAAATTTCTAATGCACGTTTGTTTTGTTTCCACAGGACGCC

>h10
GCGTAGGTGCCGCAATTGGAAAACGCATTCCCTTGGAAATAACGTACAGCTTTTTAGTCAGGAGGAACTTGATACCTAAA
TGGAAACCGTACATGGCACGACTTATGGCCAGTAAGTATTGCATATTATTAAGAACTGATCTCAATATTTTGCGTTGATG
GAGTTATAAACATACATCCGGATGCGCAGAATTGGGATAAATCCGAAATTACGTTGTAAAAAGAATAACTATACTTCACA
GATTATTTCATCAGTAACATAAAAACTGTACACATCGTTCACCGCAATAATGAGGCACAAACCACGCCCACTGTATTGTC
AGCCTGTAAGTAACACAGTGGTCCCTAGTTGGCAAGGCCTTAGACGGTCACCAGGGAACGAAATCAGTTTTGTTTCCTGT
AAAGGTATCCCGAAAGCAAATAGAATGTCAAACATCTATGGTTTAATGCTAATTTGCTAAGTTATCTCTATTTGGACTAG
TAGTGTGTCTCTGAAATCACTTTTATTTTATTTATCTATATATTTATTAAGAGCTAAGAAAATGGAACTGGTCGTCATGC
CTGGTGACTTGCCTGATTGCCCCGCAGACATGGATCGTTGTAAAAATTATAAGCTGACTCCAGACTTATTTACATGTCGC
CTTCATATAGCTGGAACTTTTGGTGAATGTCGCGTAACAAAACACTTTATTCGTAGTTTTAAGAGGTCATCCCTTCATTT
GAAGTGTATTGTTTGTTTTTTAAAAATTTCTAATGCACGTTTGTTTTGTTTCCACAGGACGCC

>h11
GCGTAGGTGCCGCAATTGGAAAACGCATTCCCTTGGAAATAACGTACAGCTTTTTAGTCAGGAGGAACTTGATACCTAAA
TGGAAACCGTACATGGCACGACTTATGGCCAGTAAGTATTGCATATTATTAAGAACTGATCTCAATATTTTGCGTTGATG
GAGTTATAAACATACATCCGGATGCGCAGAATTGGGATAAATCCGAAATTACGTTGTAAAAAGAATAACTATACTTCACA
GATTATTTCATCAGTAACATAAAAACTGTACACATCGTTCACCGCAATAATGAGGCACAAACCACGCCCACTGTATTGTC
AGCCTGTAAGTAACACAGTGGTTCCTAGTTGGCAAGGCCTTAGACGGTCACCAGGGAACGAAATCAGTTTTGTTTCCTGT
TAAGGTATCCCGAAAGCAAATAGAATGTCAAACATCTATGGTTTAATGCTAATTTGCTAAGTTATCTCTATTTGGACTAG
TAGTGTGTCTCTGAAATCACTTTTATTTTATTTATCTATATATTTATTAAGAGCTAAGAAAATGGAACTGGTCGTCATGC
CTGGTGACTTGCCTGATTGCCCCGCAGACATGGATCGTTGTAAAAATTATAAGCTGACTCCAGACTTATTTACATGTCGC
CTTCATATAGCTGGAACTTTTGGTGAATGTCGCGTAACAAAACACTTTATTCGTAGTTTTAAGAGGTCATCCCTTCATTT
GAAGTGTATTGTTTGTTTTTTAAAAATTTCTAATGCACGTTTGTTTTGTTTCCACAGGACGCC

>h12
GCGTAGGTGCCGCAATTGGAAAACGCATTCCCTTGGAAATAACGTACAGCTTTTTAGTCAGGAGGAACTTGATACCTAAA
TGGAAACCGTACATGGCACGACTTATGGCCAGTAAGTATTGCATATTATTAAGAACTGATCTCAATATTTTGCGTTGATG
GAGTTATAAACATACATCCGGATGCGCAGAATTGGGATAAATCCGAAATTACGTTGTAAAAAGAATAACTATACTTCACA
GATTATTTCATCAGTAACATAAAAACTGTACATATCGTTCACCGCAATAATGAGGCACAAACCACGCCCACTGTATTGTC
AGCCTGTAAGTAACACAGTGGTCCCTAGTTGGCAAGGCCTTAGACGGTCACCAGGGAACGAAATCAGTTTTGTTTCCTGT
TAAGGTATCCCGAAAGCAAATAGAATGTCAAACATCTATGGTTTAATGCTAATTTGCTAAGTTATCTCTATTTGGACTAG
TAGTGTGTCTCTGAAATCACTTTTATTTTATTTATCTATATATTTATTAAGAGCTAAGAAAATGGAACTGGTCGTCATGC
CTGGTGACTTGCCTGATTGCCCCGCAGACATGGATCGTTGTAAAAATTATAAGCTGACTCCAGACTTATTTACATGTCGC
CTTCATATAGCTGGAACTTTTGGTGAATGTCGCGTAACAAAACACTTTATTCGTAGTTTTAAGAGGTCATCCCTTCATTT
GAAGTGTATTGTTTGTTTTTTAAAAATTTCTAATGCACGTTTGTTTTGTTTCCACAGGACGCC

>h13
GCGTAGGTGCCGCAATTGGAAAACGCATTCCCTTGGAAATAACGTACAGCTTTTTAGTCAGGAGGAACTTGATACCTAAA
TGGAAACCGTACATGGCACGACTTATGGCCAGTAAGTATTGCATATTATTAAGAACTGATCTCAATATTTTGCGTTGATG
GAGTTATAAACATACATCCGGATGCGCAGAATTGGGATAAATCCGAAATTACGTTGTAAAAAGAATAACTATACTTCACA
GATTATTT----------CATCGTTCACCGCAATAATGAGGCACAAACCACGCCCACTGTATTGTC
AGCCTGTAAGTAACACAGTGGTCCCTAGTTGGCAAGGCCTTAGACGGTCACCAGGGAACGAAATCAGTTTTGTTTCCTGT
TAAGGTATCCCGAAAGCAAATAGAATGTCAAACATCTATGGTTTAATGCTAATTTGCTAAGTTATCTCTATTTGGACTAG
TAGTGTGTCTCTGAAATCACTTTTATTTTATTTATCTATATATTTATTAAGAGCTAAGAAAATGGA----------
------ACTTGCCTGATTGCCCCGCAGACATGGATCGTTGTAAAAATTATAAGCTGACTCCAGACTTATTTACATGTCGC
CTTCATATAGCTGGAACTTTTGGTGAATGTCGCGTAACAAAACACTTTATTCGTAGTTTTAAGAGGTCATCCCTTCATTT
GAAGTGTATTGTTTGTTTTTTAAAAATTTCTAATGCACGTTTGTTTTGTTTCCACAGGACGCC

>h14
GCGTAGGTGCCGCAATTGGAAAACGCATTCCCTTGGAAATAACGTACAGCTTTTTAGTCAGGAGGAACTTGATACCTAAA
TGGAAACCGTACATGGCACGACTTATGGCCAGTAAGTATTGCATATTATTAAGAACTGATCTCAATATTTTGCGTTGATG
GAGTTATAAACATACATCCGGATGCGCAGAATTGGGATAAATCCGAAATTACGTTGTAAAAAGAATAACTATACTTAACA
GATTATTTCATCAGTAACATAAAAACTGTACACATCGTTCACCGCAATAATGAGGCACAAACCACGCCCACTGTATTGTC
AGCCTGTAAGTAACACAGTGGTCCCTAGTTGGCAAGGCCTTAGACGGTCACCAGGGAACGAAATCAGTTTTGTTTCCTGT
TAAGGTATCCCGAAAGCAAATAGAATGTCAAACATCTATGGTTTAATGCTAATTTGCTAAGTTATCTCTATTTGGACTAG
TAGTGTGTCTCTGAAATCACTTTTATTTTATTTATCTATATATTTATTAAGAGCTAAGAAAATGGAACTGGTCGTCATGC

CTGGTGACTTGCCTGATTGCCCCGCAGACATGGATCGTTGTAAAAATTATAAGCTGACTCCAGACTTATTTACATGTCGC
CTTCATATAGCTGGAACTTTTGGTGAATGTCGCGTAACAAAACACTTTATTCGTAGTTTTAAGAGGTCATCCCTTCATTT
GAAGTGTATTGTTTGTTTTTTAAAAATTTCTAATGCACGTTTGTTTTGTTTCCACAGGACGCC

>h15
GCGTAGGTGCCGCAATTGGAAAACGCATTCCCTTGGAAATAACGTACAGCTTTTTAGTCAGGAGGAACTTGATACCTAAA
TGGAAACCGTACATGGCACGACTTATGGCCAGTAAGTATTGCATATTATTAAGAACTGATCTCAATATTTTGCGTTGATG
GAGTTATAAACATACATCCGGATGCGCAGAATTGGTATAAATCCGAAATTACGTTGTAAAAAGAATAACTATACTTCACA
GATTATTTCATCAGTAACATAAAAACTGTACACATCGTTCACCGCAATAATGAGGCACAAACCACGCCCACTGTATTGTC
AGCCTGTAAGTAACACAGTGGTCCCTAGTTGGCAAGGCCTTAGACGGTCACCAGGGAACGAAATCAGTTTTGTTTCCTGT
TAAGGTATCCCGAAAGCAAATAGAATGTCAAACATCTATGGTTTAATGCTAATTTGCTAAGTTATCTCTATTTGGACTAG
TAGTGTGTCTCTGAAATCACTTTTATTTTATTTATCTATATATTTATTAAGAGCTAAGAAAATGGAACTGGTCGTCATGC
CTGGTGACTTGCCTGATTGCCCCGCAGACATGGATCGTTGTAAAAATTATAAGCTGACTCCAGACTTATTTACATGTCGC
CTTCATATAGCTGGAACTTTTGGTGAATGTCGCGTAACAAAACACTTTATTCGTAGTTTTAAGAGGTCATCCCTTCATTT
GAAGTGTATTGTTTGTTTTTTAAAAATTTCTAATGCACGTTTGTTTTGTTTCCACAGGACGCC

>h16
GCGTAGGTGCCGCAATTGGAAAACGCATTCCCTTGGAAATAACGTACAGCTTTTTAGTCAGGAGGAACTTGATACCTAAA
TGGAAACCGTACATGGCACGACTTATGGCCAGTAAGTATTGCATATTATTAAGAACTGATCTCAATATTTTGCGTTTATG
GAGGTATAAACATACATCCGGATGCGCAGAATTGGGATAAATCCGAAATTACGTTGTAAAAAGAATAACTATACTTCACA
GATTATTTCATCAGTAACATAAAAACTGTACACATCGTTCACCGCAATAATGAGGCACAAACCACGCCCACTGTATTGTC
AGCCTGTAAGTAACACAGTGGTCCCTAGTTGGCAAGGCCTTAGACGGTCACCAGGGAACGAAATCAGTTTTGTTTCCTGT
TAAGGTATCCCGAAAGCAAATAGAATGTCAAACATCTATGGTTTAATGCTAATTTGCTAAGTTATCTCTATTTGGACTAG
TAGTGTGTCTCTGAAATCACTTTTATTTTATTTATCTATATATTTATTAAGAGCTAAGAAAATGGAACTGGTCGTCATGC
CTGGTGACTTGCCTGATTGCCCCGCAGACATGGATCGTTGTAAAAATTATAAGCTGACTCCAGACTTATTTACATGTCGC
CTTCATATAGCTGGAACTTTTGGTGAATGTCGCGTAACAAAACACTTTATTCGTAGTTTTAAGAGGTCATCCCTTCATTT
GAAGTGTATTGTTTGTTTTTTAAAAATTTCTAATGCACGTTTGTTTTGTTTCCACAGGACGCC

>h17
GCGTAGGTGCCGCAATTGGAAAACGCATTCCCTTGGAAATAACGTACAGCTTTTTAGTCAGGAGGAACTTGATACCTAAA
TGGAAACCGTACATGGCACGACTTATGGCCAGTAAGTATTGCATATTATTAAGAACTGATCTCAATATTTTGCGTTTATG
GAGGTATAAACATACATCCGGATGCGCAGAATTGGGATAAATCCGAAATTACGTTGTAAAAAGAATAACTATACTTCACA
GATTATTTCATCAGTAACATAAAAACTGTACACATCGTTCACCGCAATAATGAGGCACAAACCACGCCCACTGTATTGTC
AGCCTGTAAGTAACACAGTGGTCCCTAGTTGGCAAGGCCTTAGACGGTCACCAGGGAACGAAATCAGTTTTGTTTCCTGT
TAAGGTATCCCGAAAGCAAATAGAATGTCAAACATCTATGGTTTAATGCTAATTTGCTAAGTTATCTCTATTTGGACTAG
TAGTGTGTCTCTGAAATCACTTTTATTTTATTTATCTATATATTTATTAAGAGCTAAGAAAATGGAACTGGTCGTCATGC
CTGGTGACTTGCCTGATTGCCCCGCAGACATGGATCGTTGTAAAAATTATAAGCTGACTCCAGACTTATTTACATGTCGC
CTTCATATAGCTGGAACTTTTGGTGAATGTCGCGTAACAAAACACTTTATTCGTAGTTTTAAGAGGTCATCCCTTCATTT
GAAGTGTATTGTTTGTTTTTTAAAAATTTCTAATGCACGTTTGTTTTGTTTCCACAGGAAGCC

>h18
GCGTAGGTGCCGCAATTGGAAAACGCATTCCCTTGGAAATAACGTACAGCTTTTTAGTCAGGAGGAACTTGATACCTAAA
TGGAAACCGTACATGGCACGACTTATGGCCAGTAAGTATTGCATATTATTAAGAACTGATCTCAATATTTTGCGTTTATG
GAGGTATAAACATACATCCGGATGCGCAGAATTGGGATAAATCCGAAATTACGTTGTAAAAAGAATAACTATACTTCACA
GATTATTTCATCAGTAACATAAAAACTGTACACATCGTTCACCGCAATAATGAGGCACAAACCACGCCCACTGTATTGTC
AGCCTGTAAGTAACACAGTGGTCCCTAGTTGGCAAGGCCTTAGACGGTCACCAGGGAACGAAATCAGTTTTGTTTCCTGT
TAAGGTATCCCGAAAGCAAATAGAATGTCAAACATCTATGGTTTAATGCTAATTTGCTAAGTTATCTCTATTTGGACTAG
TAGTGTGTCTCTGAAATCACTTTTATTTTATTTATCTATATATTTATTAAGAGCTAAGAAAATGGAACTGGTCGTCATGC
CTGGTGACTTGCCTGATTGCCCCGCAGACATGGATCGTTGTAAAAATTATAAGCTGACTCCAGACTTATTTACATGTCGC
CTTCATATAGCTGGAACTTTTGGTGAATGTCGCGTAACAAAACACTTTATTCGTAGTTTTAAGAGGTCATCCCTTCATTT
GAAGTGTATTGTTTGTTTTT-AAAAATTTCTAATGCACGTTTGTTTTGTTTCCACAGGACGCC

>h19
GCGTAGGTGCCGCAATTGGAAAACGCATTCCCTTGGAAATAACGTACAGCTTTTTAGTCAGGAGGAACTTGATACCTAAA
TGGAAACCGTACATGGCACGACTTATGGCCAGTAAGTATTGCATATTATTAAGAACTGATCTCAATATTTTGCGTTTATG
GAGGTATAAACATACATCCGGATGCGCAGAATTGGGATAAATCCGAAATTACGTTGTAAAAAGAATAACTATACTTCACA
GATTATTTCATCAGTAACATAAAAACTGTACACATCGTTCACCGCAATAATGAGGCACAAACCACGCCCACTGTATTGTC

AGCCTGTAAGTAACACAGTGGTCCCTAGTTGGCAAGGCCTTAGACGGTCACCAGGGAACGAAATCAGTTTTGTTTCCTGT
TAAGGTATCCCGAAAGCAAATAGAATGTCAAACATCTATGGTTTAATGCTAATTTGCTAAGTTATCTCTATTTGGACTAG
TAGTGTGTCTCTGAAATCACTTTTATTTTATTTATCTATATATTTATTAAGAGCTAAGAAAATGGAACTGGTCGTCATGC
CTGGTGACTTGCCTGATTGCCCCGCAGACATGGATCGTTGTAAAAATTATAAGCTGACTCCAGACTTATTTACATGTCGC
CTTCATATAGCTGGAACTTTGGGTGAATGTCGCGTAACAAAACACTTTATTCGTAGTTTTAAGAGGTCATCCCTTCATTT
GAAGTGTATTGTTTGTTTTTTAAAAATTTCTAATGCACGTTTGTTTTGTTTCCACAGGACGCC

>h20
GCGTAGGTGCCGCAATTGGAAAACGCATTCCCTTGGAAATAACGTACAGCTTTTTAGTCAGGAGGAACTTGATACCTAAA
TGGAAACCGTACATGGCACGACTTATGGCCAGTAAGTATTGCATATTATTAAGAACTGATCTCAATATTTTGCGTTTATG
GAGGTATAAACATACATCCGGATGCGCAGAATTGGGATAAATCCGAAATTACGTTGTAAAAAGAATAACTATACTTCACA
GATTATTTCATCAGTAACATAAAAACTGTACACATCGTTCACCGCAATAATGAGGCACAAACCACGCCCACTGTATTGTC
AGCCTGTAAGTAACACAGTGGTCCCTAGTTGGCAAGGCCTTAGACGGTCACCAGGGAACGAAATCAGTTTTGTTTCCTGT
TAAGGTATCCCGAAAGCAAATAGAATGTCAAACATCTATGGTTTAATGCTAATTTGCTAAGTTATCTCTATTTGGACTAG
TAGTGTGTCTCTGAAATCACTTTTATTTTATTTATCTATATATTTATTAAGAGCTAAGAAAATGGAACTGGTCGTCATGC
CTGGTGACTTGCCTGATTGCCCCGCAGACATGGATCGTTGTAAAAATTATAAGCTGACTCCAGACTTATTTACATGTCGC
CTTCATAAAGCTGGAACTTTTGGTGAATGTCGCGTAACAAAACACTTTATTCGTAGTTTTAAGAGGTCATCCCTTCATTT
GAAGTGTATTGTTTGTTTTTTAAAAATTTCTAATGCACGTTTGTTTTGTTTCCACAGGACGCC

>h21
GCGTAGGTGCCGCAATTGGAAAACGCATTCCCTTGGAAATAACGTACAGCTTTTTAGTCAGGAGGAACTTGATACCTAAA
TGGAAACCGTACATGGCACGACTTATGGCCAGTAAGTATTGCATATTATTAAGAACTGATCTCAATATTTTGCGTTTATG
GAGGTATAAACATACATCCGGATGCGCAGAATTGGGATAAATCCGAAATTACGTTGTAAAAAGAATAACTATACTTCACA
GATTATTTCATCAGTAACATAAAAACTGTACACATCGTTCACCGCAATAATGAGGCACAAACCACGCCCACTGTATTGTC
AGCCTGTAAGTAACACAGTGGTCCCTAGTTGGCAAGGCCTTAGACGGTCACCAGGGAACGAAATCAGTTTTGTTTCCTGT
TAAGGTATCCCGAAAGCAAATAGAATGTCAAACATCTATGGTTTAATGCTAATTTGCTAAGTTATCTCTATTTGGACTAG
TAGTGTGTCTCTGAAATCACTTTTATTTTATTTATCTATATATTTATTAAGAGCTAAGAAAATGGAACTGGTCGTCATGC
CTGGCGACTTGCCTGATTGCCCCGCAGACATGGATCGTTGTAAAAATTATAAGCTGACTCCAGACTTATTTACATGTCGC
CTTCATATAGCTGGAACTTTTGGTGAATGTCGCGTAACAAAACACTTTATTCGTAGTTTTAAGAGGTCATCCCTTCATTT
GAAGTGTATTGTTTGTTTTTTAAAAATTTCTAATGCACGTTTGTTTTGTTTCCACAGGACGCC

>h22
GCGTAGGTGCCGCAATTGGAAAACGCATTCCCTTGGAAATAACGTACAGCTTTTTAGTCAGGAGGAACTTGATACCTAAA
TGGAAACCGTACATGGCACGACTTATGGCCAGTAAGTATTGCATATTATTAAGAACTGATCTCAATATTTTGCGTTTATG
GAGGTATAAACATACATCCGGATGCGCAGAATTGGGATAAATCCGAAATTACGTTGTAAAAAGAATAACTATACTTCACA
GATTATTTCATCAGTAACATAAAAACTGTACACATCGTTCACCGCAATAATGAGGCACAAACCACGCCCACTGTATTGTC
AGCCTGTAAGTAACACAGTGGTCCCTAGTTGGTGAGGCCTTAGACGGTCACCAGGGAACGAAATCAGTTTTGTTTCCTGT
TAAGGTATCCCGAAAGCAAATAGAATGTCAAACATCTATGGTTTAATGCTAATTTGCTAAGTTATCTCTATTTGGACTAG
TAGTGTGTCTCTGAAATCACTTTTATTTTATTTATCTATATATTTATTAAGAGCTAAGAAAATGGAACTGGTCGTCATGC
CTGGTGACTTGCCTGATTGCCCCGCAGACATGGATCGTTGTAAAAATTATAAGCTGACTCCAGACTTATTTACATGTCGC
CTTCATATAGCTGGAACTTTTGGTGAATGTCGCGTAACAAAACACTTTATTCGTAGTTTTAAGAGGTCATCCCTTCATTT
GAAGTGTATTGTTTGTTTTTTAAAAATTTCTAATGCACGTTTGTTTTGTTTCCACAGGACGCC

>h23
GCGTAGGTGCCGCAATTGGAAAACGCATTCCCTTGGAAATAACGTACAGCTTTTTAGTCAGGAGGAACTTGATACCTAAA
TGGAAACCGTACATGGCACGACTTATGGCCAGTAAGTATTGCATATTATTAAGAACTGATCTCAATATTTTGCGTTTATG
GAGGTATAAACATACATCCGGATGCGCAGAATTGGGATAAATCCGAAATTACGTTGTAAAAAGAATAACTATACTTCACA
GATTATTTCATCAGTAACAAAAAAACTGTACACATCGTTCACCGCAATAATGAGGCACAAACCACGCCCACTGTATTGTC
AGCCTGTAAGTAACACAGTGGTCCCTAGTTGGCAAGGCCTTAGACGGTCACCAGGGAACGAAATCAGTTTTGTTTCCTGT
TAAGGTATCCCGAAAGCAAATAGAATGTCAAACATCTATGGTTTAATGCTAATTTGCTAAGTTATCTCTATTTGGACTAG
TAGTGTGTCTCTGAAATCACTTTTATTTTATTTATCTATATATTTATTAAGAGCTAAGAAAATGGAACTGGTCGTCATGC
CTGGTGACTTGCCTGATTGCCCCGCAGACATGGATCGTTGTAAAAATTATAAGCTGACTCCAGACTTATTTACATGTCGC
CTTCATATAGCTGGAACTTTTGGTGAATGTCGCGTAACAAAACACTTTATTCGTAGTTTTAAGAGGTCATCCCTTCATTT
GAAGTGTATTGTTTGTTTTTTAAAAATTTCTAATGCACGTTTGTTTTGTTTCCACAGGACGCC

>h24
GCGTAGGTGCCGCAATTGGAAAACGCATTCCCTTGGAAATAACGTACAGCTTTTTAGTCAGGAGGAACTTGATACCTAAA

TGGAAACCGTACATGGCACGACTTATGGCCAGTAAGTATTGCATATTATTAAGAACTGATCTCAATATTTTGCGTTTATG
GAAGTATAAACATACATCCGGATGCGCAGAATTGGGATAAATCCGAAATTACGTTGTAAAAAGAATAACTATACTTCACA
GATTATTTCATCAGTAACATAAAAACTGTACACATCGTTCACCGCAATAATGAGGCACAAACCACGCCCACTGTATTGTC
AGCCTGTAAGTAACACAGTGGTCCCTAGTTGGCAAGGCCTTAGACGGTCACCAGGGAACGAAATCAGTTTTGTTTCCTGT
TAAGGTATCCCGAAAGCAAATAGAATGTCAAACATCTATGGTTTAATGCTAATTTGCTAAGTTATCTCTATTTGGACTAG
TAGTGTGTCTCTGAAATCACTTTTATTTTATTTATCTATATATTTATTAAGAGCTAAGAAAATGGAACTGGTCGTCATGC
CTGGTGACTTGCCTGATTGCCCCGCAGACATGGATCGTTGTAAAAATTATAAGCTGACTCCAGACTTATTTACATGTCGC
CTTCATATAGCTGGAACTTTTGGTGAATGTCGCGTAACAAAACACTTTATTCGTAGTTTTAAGAGGTCATCCCTTCATTT
GAAGTGTATTGTTTGTTTTTTAAAAATTTCTAATGCACGTTTGTTTTGTTTCCACAGGACGCC

APPENDIX 9: LYSIN GENOTYPE FREQUENCIES

Hap ID	CBL	CCB	CRW	DSD	GLN	GOB	IHM	JCH	MAT	MTB	NPT	OCH	SPB	STR	TCL	TIM	WST	Total
h1, h1	7	13	5	10	10	12	10	9	7	17	11	9	9	2	13	12	4	160
h1, h16	4	5	4	4	7	2	3	8	6	4	7	3	6	3	5	6	5	82
h1, h4	1		2															3
h1, h24	1												1					2
h16, h16	1				1		1	1		1			1			2		8
h1, h5			1															1
h1, h22			1															1
h1, h9			2				1											3
h6, h16			1															1
h1, h12			1													1		2
h16, h21			1															1
h1, h15					1								2					3
h1, h10					1													1
h1, h13						1					1					1		3
h4, h6							1											1
h1, h20							1											1
h1, h8							1											1
h16, h19							1											1
h1, h18							1								1			2
h1, h2								1								1		2
h1, h3								1										1
h1, h6									1									1
h1, h21									2									2
h1 ,h7									1									1
h1, h23									1									1
h1, h14											1							1
h1, h17													1					1
h16, h24														1				1
h1, h11															1			1
Total	14	18	18	14	20	15	20	20	18	22	20	12	20	6	20	23	9	289

APPENDIX 10: Gα1 INTRON HAPLOTYPES

Listed below is the alignment in FASTA format for the 112 Gα1 haplotypes. Labels refer to the numbered haplotypes in Figure 6.2. Note, there is no h65.

```
>h1
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCACTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
AAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCATCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h2
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCACTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
AAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAA-CCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATA----------TAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h3
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h4
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
```

TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAAAAGGTA

>h5
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h6
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCTGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h7
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCTGCGTCATGCCAAAAAACGTTTAAAAGATGGTA

>h8
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAATATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h9
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG

TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAATATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACATTTAAAAGATGGTA

>h10
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAATATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h11
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATTTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h12
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACA----------TGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGTGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h13
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA

ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h14
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCGTGCCAAAAGACGTTTAAAAGATGGTA

>h15
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATAGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h16
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATAGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAATAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h17
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCACTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h18
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCACTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTATAAAAGATGGTA

>h19
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGAAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h20
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACATCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCTGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h21
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGATCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATTTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h22
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAATTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACAATTCACCCGAGCAAATCCCA

GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGCCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h23
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAATTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACAATTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTT--TGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTGACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h24
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAATTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACATTAATTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGTGTTCGATCCCCGGCCGCGTCATACCAAAAGACGTTTAAAAGATGGTA

>h25
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGAGTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCCCCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCTGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h26
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGAGTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCCCCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAA-CCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT

TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h27
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGCACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h28
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGCACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGCCCGCGTCATGCCAAAAAACGTTTAAAAGATGGTA

>h29
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGCACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCAGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h30
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAGCATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h31
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-

----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAGCATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h32
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAGCATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAAAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h33
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAGCATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACAAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h34
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAGCATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGATCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h35
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCTTATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGAGTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCCCCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG

ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAGAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCTGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h36
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGAATTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAATTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAAC----------------ATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h37
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTTATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h38
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACATGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCTGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h39
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACATGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAATATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACATTTATAAGATGGTA

>h40
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACATGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h41
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTATGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h42
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h43
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGCCATGCCAAAAGACGTTTAAAAGATGGTA

>h44
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC

CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCATCATGCCAAAAGACGTTTAAAAGATGGTA

>h45
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAAAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h46
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTTTTACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h47
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h48
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG

CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGA-------
-TATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h49
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAAAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h50
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAATTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACAATTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h51
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAATTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACAATTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGTCGT--
----AAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTGT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAAGAA-CCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATA----------TAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h52
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAATTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACAATTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGTCGT--
----AAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTGT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAAGAA-CCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATA----------TAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAAAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h53

AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAATTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACATTAATTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGTGTTCGATCCCCGGCCGCGTCATACCAAAAGACGTTTAAAAGATGGTA

>h54
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAATTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACATTAATTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGATGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATACCAAAAGACGTTTAAAAGATGGTA

>h55
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAATTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACATTAATTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCAGGACC-GGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGATGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATACCAAAAGACGTTTAAAAGATGGTA

>h56
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTAATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h57
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGCACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA

AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h58
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGCACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTTGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h59
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGCACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAATTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h60
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGCACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAATTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAATAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h61
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGCACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAATTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACAATTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTTTTACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG

TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h62
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGCACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAATTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACAATTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h63
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGCACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAATTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACAATTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGA-------
-TATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h64
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGCACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAATTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACAATTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGA-------
-TATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACATTTAAAAGATGGTA

>h66
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGCACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAATTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACAATTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGGCCCGGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGA-------
-TATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h67
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG

TGCCATATTTTAACATAGGCACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAATTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACAATTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCAACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-AGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h68
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGCACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAATTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACAATTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCAACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-AGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h69
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGCACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAATTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACAATTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTT--TGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h70
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGCACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCGATTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACAATTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCAACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-AGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h71
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGCACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTAATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA

ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h72
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTTCTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h73
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACATTACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h74
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACCGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h75
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTATGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h76
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTATGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
CTATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h77
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTATGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTACCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGATAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h78
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTATGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h79
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTATGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGCACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAATTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTACCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGATAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h80
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTATGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGCACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAATTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACAATTCACCCGAGCAAATCCCA

GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h81
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTATGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGCACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTGATAGCTGTC
CAACTGTCCATTAGCCAATTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAATAATTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h82
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTATGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGCGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h83
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTATGAGACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h84
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGACAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGCACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGGCTGGACC-GGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCAGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT

TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h85
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAATCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGAAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h86
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAATCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGAAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGTTATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h87
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTTGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCACTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h88
AAGGAGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTAATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTATGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h89
AAGGAGCTTGTGTTTCTCTGGCAAGAACATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-

----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGTGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCTGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h90
AAGGAGCTTATGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCACTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGCACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAATTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACAATTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h91
AAGGAGCTTATGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h92
AAGGAGCTTATGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h93
AAGGAGCTTATGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG

ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h94
AAGGAGCTTATGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h95
AAGGAGCTTATGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACATCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h96
AAGGAGCTTATGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAATTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACAATTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGTCGT--
----AAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTGT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAAGAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCATCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h97
AAGGAGCTTATGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATCGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGGTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GAACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h98
AAGGAGCTTATGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGCACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAATTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACAATTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATACCAAAAGACGTTTAAAAGATGGTA

>h99
AAGGAGCTTATGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACATTACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAGGACGTTTAAAAGATGGTA

>h100
AAGGAGCTTATGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGGTTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACATCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h101
AAGGAGCTTATGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h102
AAGGAGCTTATGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGCACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC

CAACTGTCCATTAGCCAATTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACAATTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTCGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGA-------
-TATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h103
AAGGAGCTTATGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGCCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h104
AAGGAGCTTATGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGCCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCATCATGCCAAAAGACGTTTAAAAGATGGTA

>h105
AAGGAGCTTATGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGCCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCATCATGCCAAAAGACGTTTAAAAGATCGTA

>h106
AAGGAGCTTATGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGCCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG

CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCATCATGCCAAAAGACGTTTAAAAAATGGTA

>h107
AAGGAGCTTATGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGCCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCATCATGCCAAAAAACGTTTAAAAGATGGTA

>h108
AAGGAGCTTATGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGCCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCGGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCATCATGCCAAAAGACGTTTAAAAGATGGTA

>h109
AAGGAGCTTATGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTATGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h110
AAGGAGCTTATGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGGTTACGAAACAGCCAGTCTGTGTTAAGACAATACACGCCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCATCATGCCAAAAGACGTTTAAAAGATGGTA

>h111

AAGGAGCTTATGTTTCTCTGGCAAGACCATTAATTCTATTAATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGTCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h112
AAGGGGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAA-
----AGGTAGATTACGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAATATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

>h113
AAGGGGCTTGTGTTTCTCTGGCAAGACCATTAATTCTATTTATTGTTCAATGAATTTCTTCGCTTGCAGTCAAGGTAAAG
GAGAAGGTAGATTATGAAACAGCCAGTCTGTGTTAAGACAATACACGTCTGAGAGTACTATTTATCTAGTGATTTCAGGG
TGCCATATTTTAACATAGGTACATGATAGCCTTGGCGCTATGATTGCTTTGTGGAATGGAAACCTAGGCTCATAGCTGTC
CAACTGTCCATTAGCCAGTTGTCCTGAATTTGTTGTTTGTTTTGAATTTATAAAACAACATTTCACCCGAGCAAATCCCA
GCATCACCATGGTCACCGTCCAAACTAGACCCATGAAGACCTTGGTAGAATACGTCTTCAGCAACCCATGCTTGATGTAA
AAGGAAGAGGCGACTAAGGGGATCGGGTGGTCAGACTCGCTAACTTGGTTGATGCATGTCATCGTATCCCAATTGCATGG
ATTGATTCTCATGTTTTCAATCACTGGATTGTCTGGACC-GGACTTGATTTTTTT-ACAGACTGCCATCACATAGCTGGA
ATATTGCTGAGTGCGGCGTTAAACAACAAACCAACACAACACATCGCCCAAACAATAGGGTCATGCACTGAAATCTCATG
CTTTGCGACCTAACACCTAGTAACTCGGAAATATATAAATTGGTGTAACATGTTCTGTTTTGAAAAATGATGACAATTTT
TTATGGATATAAAACTTCTTTATACTGTGAAGGTGACGTTTCAACATAGATTCTAATGTCGAAATAGAATAAAGAAGTTG
TATATCGTGGGTTCGATCCCCGGCCGCGTCATGCCAAAAGACGTTTAAAAGATGGTA

APPENDIX 11: Gα1 INTRON GENOTYPE FREQUENCIES

Genotype	CCB	CRW	DSD	GLN	GOB	IHM	MAT	NPT	OCH	SPB	STR	TCL	TIM	WST	Total
h1, h6	1														1
h92, h92	2							1				1			4
h37, h45	1														1
h75, h75	2				2			2					1		7
h54, h86	1														1
h92, h103	1														1
h47, h75	1							1				1			3
h59, h92	1										1				2
h75, h77	1			1				1							3
h6, h79	1														1
h112, h113	1														1
h6, h78	1														1
h77, h104	1														1
h5, h104	1														1
h92, h109	1														1
h15, h92		2													2
h29, h42		1													1
h6, h62		1													1
h82, h86		1													1
h61, h94		1													1
h15, h94		1													1
h35, h91		1													1
h24, h66		1													1
h27, h27		1													1
h6, h47		1													1
h10, h26		1													1
h27, h75		1	1												2
h24, h59			1												1
h77, h86			1												1
h6, h75			1		1		1	1				1			5
h37, h63			1										2		3
h75, h104			1		1								2		4
h54, h75			1					1							2
h42, h77			1												1
h8, h86			1		1										2
h69, h69			1												1
h100, h110			1												1
h62, h68				1						1					2
h3, h31				1											1
h13, h68				1											1
h27, h30				1											1
h38, h86				1											1
h6, h13				1											1
h13, h93				1											1
h68, h75				1						1					2
h15, h38				1											1
h4, h106				1											1
h22, h56				1											1
h3, h23				1											1
h47, h71				1											1
h31, h94				1											1
h11, h31				1											1
h11, h49				1											1
h13, h31				1											1
h62, h93				1											1
h11, h95					1										1
h5, h101					1										1
h8, h63					1			1						1	3
h69, h75					1										1
h75, h92					1						1	1	1		4
h41, h57					1										1

h47, h92					1										1
h6, h8					1										1
h6, h37					1										1
h73, h99					1										1
h86, h92					1										1
h75, h86					1			2							3
h24, h104					1										1
h50, h56						1									1
h3, h75						1		1							2
h24, h84						1									1
h71, h89						1									1
h6, h104						1									1
h6, h91						1									1
h13, h56						1									1
h56, h64						1									1
h37, h58						1									1
h5, h24						1									1
h69, h89						1									1
h1, h12						1									1
h11, h14						1									1
h25, h87						1									1
h56, h68						1									1
h21, h34						1									1
h2, h96						1									1
h6, h67						1									1
h8, h42						1		1							2
h11, h55						1									1
h1, h75							1					1			2
h3, h91							1								1
h16, h81							1								1
h33, h59							1								1
h38, h38							1								1
h37, h53							1								1
h68, h77							1								1
h56, h90							1								1
h1, h77							1								1
h11, h54							2								2
h24, h76							1								1
h11, h62							1								1
h11, h94							1								1
h36, h42							1								1
h75, h95								1							1
h37, h92								1							1
h86, h104								1							1
h24, h75								1							1
h42, h42								1							1
h16, h77								1							1
h77, h95								1							1
h6, h48								1							1
h25, h63									1						1
h54, h63									1				1		2
h63, h92									1						1
h50, h75									2						2
h70, h108									1						1
h68, h86									1						1
h3, h3									1						1
h63, h75									1				1	1	3
h16, h60									1						1
h47, h63									1						1
h25, h95									1						1
h40, h68									1						1
h22, h69										1					1
h24, h95										1					1
h3, h15										1					1
h18, h56										1					1
h91, h97										1					1

h51, h80										1					1
h15, h63										1					1
h3, h47										1					1
h3, h86										1					1
h32, h52										1					1
h92, h98										1					1
h47, h86										1					1
h3, h55										1					1
h9, h102										1					1
h46, h71										1					1
h42, h56										1					1
h38, h92										1					1
h22, h68										1					1
h31, h44										1					1
h91, h104											1				1
h54, h92											1		2		3
h7, h107											1				1
h104, h105											1				1
h17, h54												1			1
h6, h38												1			1
h25, h39												1			1
h43, h92												1			1
h47, h68												1			1
h85, h92												1			1
h6, h11												1			1
h6, h92												1	1		2
h1, h59												1			1
h6, h93												1			1
h5, h92												1			1
h37, h54												1			1
h68, h92												1			1
h6, h68												1			1
h72, h92												1			1
h59, h75													1		1
h8, h47													1		1
h24, h42													1		1
h42, h47													1		1
h63, h77													1		1
h6, h86													1		1
h8, h92													1		1
h62, h75													1		1
h28, h104													1		1
h37, h75													1		1
h88, h111													1		1
h8, h75													1	1	2
h24, h83														1	1
h20, h92														1	1
h8, h77														1	1
h37, h74														1	1
h38, h42														1	1
h42, h75														1	1
h6, h19														1	1
	17	13	11	19	18	20	16	20	13	21	6	20	23	10	227